# UNTERSUCHUNGEN ZU LERNPSYCHOLOGISCH SPEZIFIZIERTEN ASPEKTEN DES INTERAKTIONS- UND UNTERRICHTSSTILS VON LEHRERN AUS DER SICHT DER SCHÜLER

Dissertation
zur
Erlangung des Grades eines Doktors
der Erziehungswissenschaften
an der
Pädagogischen Hochschule Ruhr
in Dortmund

vorgelegt von
Friedrich Masendorf
Dortmund

1973

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41536502_0007

Druck- und Verlagsanstalt

Werner Blasaditsch · 89 Augsburg · Theodor-Heuss-Platz 3 · Telefon (0821) 25563

Printed in Germany

# UNTERSUCHUNGEN ZU LERNPSYCHOLOGISCH SPEZIFIZIERTEN ASPEKTEN DES INTERAKTIONS- UND UNTERRICHTSSTILS VON LEHRERN AUS DER SICHT DER SCHÜLER

DISSERTATION
zur
Erlangung des Grades eines Doktors
der Erziehungswissenschaften
an der
Pädagogischen Hochschule Ruhr
in Dortmund

vorgelegt von

Friedrich Masendorf

Dortmund

UNIV.
LEIDEN
BIBL.

Referent: Prof. Dr. A. Reinartz
Korreferent: Prof. Dr. W. Brüne
Tag der mündlichen Prüfung: 6. März 1973

Meinen Eltern gewidmet

Kurze Zusammenfassung des Inhalts der Untersuchung

Gegenstand dieser Untersuchung ist die Konstruktion und Validierung einer Erziehungsskala zur Erfassung bestimmter, lernpsychologisch spezifizierter Aspekte des Interaktions- und Unterrichtsstils von Lehrern aus der Sicht der Schüler, wodurch eine ökonomische Möglichkeit zur Analyse von schulischen Lehr- und Lernprozessen aufgezeigt und erprobt wird.

Die Brauchbarkeit der Erziehungsskala wird an Schülerpopulationen unterschiedlichen Schultyps, insbesondere auch in Lernbehindertenklassen, überprüft. Insgesamt sind an der Untersuchung 1707 Schüler und Schülerinnen aus 66 Klassen beteiligt. Die Erziehungsskala gibt Aufschluß über solche Verhaltensweisen von Lehrern, die für Klima und Lernatmosphäre in Schulklassen langfristig von Bedeutung sind.

Für die Endform der als Beobachtungsinstrument konzipierten Erziehungsskala werden solche Items berücksichtigt, die

a) dem Ergebnis eines Expertenratings zufolge in extrem positivem und extrem negativem Ausmaß zum Klassenklima beitragen;
b) die Tendenz zu sozial erwünschtem Antwortverhalten der Schüler weitgehend ausschließen;
c) in Bezug auf perzipiertes Belohnungs- und Bestrafungsverhalten von Lehrern hohe Trennschärfen aufweisen;
d) Lehrer und Lehrergruppen voneinander differenzieren und
e) für Real-, Haupt- und lernbehinderte Sonderschüler siebter, achter und neunter Schuljahre eindeutig beobachtbar und registrierbar sind.

Die Reliabilitäten werden gesondert für die oben genannten Schülerpopulationen errechnet. Sie liegen um .90 und kommen damit denen von Intelligenz- und Schulleistungstests nahe.

Weiterhin wird untersucht, inwieweit der von Lehrern erwünschte von dem durch die Schüler perzipierten Unterrichtsstil abweicht. Darüberhinaus erhalten wir Aufschlüsse darüber, mit welcher Genauigkeit Lehrer Urteile über ihren von den Schülern perzipierten Unterrichtsstil abgeben können, d.h. inwieweit sie ihr eigenes Verhalten im Urteil ihrer Schüler einzuschätzen in der Lage sind. Aus der Vielfalt der Wirkungszusammenhänge, die sich in der Schüler-Lehrer-Beziehung abspielen, stellt die Genauigkeit der Wahrnehmung des Lehrers über sein Verhalten im Urteil seiner Schüler eine wichtige Variable im Hinblick auf eine Optimierung des Klassenklimas dar.

Folgende wichtige Ergebnisse lassen sich herausstellen:

- Die Lehrer können das Ausmaß ihrer von den Schülern perzipierten Strenge (ihr Bestrafungsverhalten) besser einschätzen als ihr Ausmaß an Unterstützung (ihr Belohnungsverhalten).
- Die Lehrer unterliegen in der Einschätzung ihres Unterrichtsverhaltens im Urteil ihrer Schüler gewissermaßen einem Trend zur Mitte:
  Lehrer, die im Sinne eines guten Klassenklimas von ihren Schülern eher ungünstig eingestuft werden, überschätzen ihr Ausmaß angebrachten Interaktionsverhaltens, eher günstig eingestufte Lehrer unterschätzen es.
- Ungünstig eingestufte Lehrer weisen weit höhere Diskrepanzen zwischen ihrem Wunsch- und Realverhalten auf als günstig eingestufte Lehrer.
- Lehrer, die ihr implizites Persönlichkeitsmodell zur aktuellen Schülerbeurteilung (nach HOFER, 1969) im Sinne der Urteilsdimensionen Disziplinertheit vs. Un-

disziplinertheit, geistige Regsamkeit und Normenkonformität besonders streng realisieren, unterschätzen ihr Ausmaß angebrachten Interaktionsverhaltens; Lehrer, die ihr implizites Persönlichkeitsmodell zur aktuellen Schülerbeurteilung besonders milde realisieren, überschätzen ihr Ausmaß angebrachten Interaktionsverhaltens.

Die Befunde werden unter methodologischen und erziehungspsychologischen Gesichtspunkten diskutiert.

I. EINLEITUNG: STAND DES PROBLEMS

1. Einführung in die Fragestellungen und den Themenbereich

2. Zur Methodologie
2.1. Allgemeine Bemerkungen zur Konstruktion von Beobachtungsschemata
2.2. Erhebungsmethoden zur Erfassung des Erziehungsstils
2.2.1. Unmittelbare Beobachtung
2.2.2. Selbstbeschreibung
2.2.3. Einschätzung durch Schüler
2.2.4. Kritische Würdigung der Erhebungsmethoden
2.3. Das Konzept vom Erziehungsstil
2.4. Lerntheoretische Spezifizierung eines Erziehungsstil - Modells als theoretischer Ausgangspunkt der Untersuchung
2.5. Operationalisierung von Strenge und Unterstützung
2.6. Hinweise zum gesamten Untersuchungsverlauf

II. VORUNTERSUCHUNG

3. Fragestellungen der Voruntersuchung
4. Durchführung der Voruntersuchung
4.1. Zusammenstellung des zu analysierenden Itempools
4.2. Ablauf der Datenerhebung
4.3. Auswertung der Daten und Darstellung der Ergebnisse
4.3.1. Faktorenanalyse der Items
4.3.2. Aufgabenanalyse und Reliabilität der Subskalen

4.3.3. Varianzanalytische und korrelationsstatistische Auswertung des Zahlenmaterials der Tabelle 4 im Hinblick auf die Fragestellungen
4.4. Zusammenfassung der Ergebnisse der Voruntersuchung
4.5. Diskussion der Ergebnisse der Voruntersuchung

III. ERSTER TEIL DER HAUPTUNTERSUCHUNG

5. Vorbemerkungen zum Ablauf der Hauptuntersuchung
6. Fragestellungen der Hauptuntersuchung
7. Durchführung der Hauptuntersuchung
7.1. Untersuchungsmaterial
7.1.1. Die Items zur Lehrerbeurteilung
7.1.2. Fragebogen zur Erfassung von Simulationstendenzen
7.1.3. Die Urteilseigenschaften für die Lehrer zur aktuellen Schülerbeurteilung
7.2. Ablauf der Datenerhebung
7.3. Auswertung der Daten und Darstellung der Ergebnisse
7.3.1. Faktorenanalyse der Items
7.3.2. Aufgabenanalyse und Reliabilität für die Subskalen Strenge und Unterstützung
7.3.3. Korrelationen zwischen den Beurteilungsmodi für Strenge und Unterstützung
7.3.4. Kritik an den faktorenanalytisch gewonnenen Subskalen Strenge und Unterstützung
7.3.5. Skalenkonstruktion anhand interner und externer Selektionskriterien
7.3.5.1. Vorbemerkungen
7.3.5.2. Itemauswahl aufgrund von Expertenrating
7.3.5.3. Bestimmung der Item-Validitätskoeffizienten hinsichtlich des Kriteriums soziale Erwünschtheit

7.3.5.4. Varianzanalytische Überprüfung der Itembrauchbarkeit
7.3.5.5. Itemselektion und Zusammenstellung der Fragebogenendform

IV. ZWEITER TEIL DER HAUPTUNTERSUCHUNG

8. Validierungsaspekte der Fragebogenendform
8.1. Vorbemerkungen
8.2. Statistische Auswertung des Zahlenmaterials im Hinblick auf die Validierungsfragen
9. Zusammenfassung der Ergebnisse der Hauptuntersuchung
10. Diskussion der Ergebnisse der Hauptuntersuchung

Literaturverzeichnis

Anhang

# I. EINLEITUNG: STAND DES PROBLEMS

## Einführung in die Fragestellungen und den Themenbereich

Mit der vorliegenden Entwicklung eines Instrumentariums zur Beobachtung von Unterricht soll eine ökonomische Möglichkeit zur Analyse und Beurteilung von schulischen Lehr- und Lernprozessen aufgezeigt und erprobt werden. Die mit einem solchen Vorhaben verbundenen Schwierigkeiten sind vor allem in der Komplexität des zu beobachtenden Gegenstandes Unterricht zu sehen. Da es bislang kaum Möglichkeiten gibt, alle Merkmale eines unterrichtlichen Lehr- und Lernprozesses in ihrer Gesamtheit zu erfassen, beschränkt man sich üblicherweise darauf, den mehrdimensionalen Beobachtungsgegenstand Unterricht unter ganz bestimmten, inhaltlich klar umrissenen Gesichtspunkten zu erfassen und zu analysieren.

Gegenstand dieser empirischen Untersuchung ist die Konstruktion und Validierung einer Erziehungsskala zur Erfassung bestimmter Aspekte des Interaktions- und Unterrichtsstils von Lehrern aus der Sicht der Schüler. Die Brauchbarkeit dieser Erziehungsskala soll an Schülerpopulationen unterschiedlichen Schultyps, insbesondere auch in Lernbehindertenklassen, überprüft werden. Dabei richten wir unser Augenmerk auf solche Verhaltensweisen von Lehrern, die für Klassenklima und Lernatmosphäre in Schulklassen l a n g f r i s t i g von Bedeutung sind.

Nun sind Lernatmosphäre und Klima in einer bestimmten Schulklasse die Resultante einer vielfältigen Anzahl von Einflüssen in Verbindung mit den spezifischen Charakteristika der Schüler, des Lehrers, des Lehrplans, des Unterrichtsstoffes usw. Selbst ein Beobachter, der nur gelegentlich in Unterrichtsstunden hospitiert, bemerkt

leicht, daß verschiedene Schulklassen einen unterschiedlichen Eindruck auf ihn machen. Er erkennt mehr oder weniger deutlich, daß die Klassen hinsichtlich des Ausmaßes an Disziplin, die in ihnen herrscht, aber auch hinsichtlich ihrer Spontaneität beträchtlich variieren. Klassen unterscheiden sich in dem Grad ihrer Spannung bzw. Anspannung, die im unterrichtlichen Lehr- und Lernprozeß beobachtet werden kann, wie auch in ihrer Ängstlichkeit, ihrer Bereitschaft Verantwortung zu zeigen, ihrer Ausgeglichenheit, Freizügigkeit und Ungezwungenheit. Darüberhinaus differieren Klassen auch in dem Grad, in dem sie Lenkung, emotionale Zuwendung und Verständnis durch den Lehrer erfahren; sie differieren im Ausmaß ihrer intellektuellen Anregbarkeit und emotionalen Erregbarkeit, wie auch in ihrer Bereitschaft, Toleranz gegenüber anderen Meinungen und Ansichten zu zeigen und in der Oberflächlichkeit oder Tiefe, mit der Probleme bzw. Streitfragen diskutiert und gelöst werden.

Diese und nicht zuletzt die Verhaltensmerkmale des Lehrers ergeben gemeinsam das Klassenklima, in dem sich die Schüler einer Klasse befinden.

Bei der Suche nach zeitstabilen (intraindividuell konstanten) Verhaltensmerkmalen von Lehrern soll lediglich jenes Lehrerverhalten berücksichtigt und von Schülern registriert werden, das für Schüler beobachtbar ist bzw. das mittelbar oder unmittelbar auf die Schüler wirkt und worauf diese reagieren. Von Schülern als Beobachter kann nicht erwartet werden, aus Verbal- oder sonstigem Verhalten des Lehrers etwaige "geheime Absichten" des Lehrers zu folgern und gewissermaßen interpretierte sprachliche Äußerungen zu klassifizieren. Ein derart methodisches Vorgehen führt zur mangelnden Objektivität bei der Klassifikation, da die Registrierung abhängig ist vom subjektiven Urteil des Einzelnen über meist unbeobachtbare Sachverhalte (Absichten im Lehrerverhalten). Der Grad der Objektivität eines Beobachtungsinstrumentes hängt mithin davon ab, wie "leicht" einem

Beobachter die Einstufung des beobachtbaren Verhaltens gemacht wird. Deshalb sollen solche beobachtbaren Verhaltensmerkmale gefunden werden, die neben einem hohen Informationswert über die angestellten Beobachtungen eine große Trennschärfe besitzen. Für die Endform des Beobachtungsinstrumentariums sollen insbesondere solche Verhaltensmerkmale berücksichtigt werden, die von Schülern aller Schultypen, insbesondere auch von Schülern in Lernbehindertenklassen, klar und eindeutig beobachtbar und leicht registrierbar sind und quantitative Vergleiche zwischen Lehrern bzw. Lehrergruppen zulassen.

Als weiteres Ziel soll diese Arbeit Möglichkeiten aufzeigen, die Lehrer-Schüler-Interaktion zu optimalisieren. Einen Ansatzpunkt hierzu sehen wir - ganz im Sinne des sozialpsychologischen Kommunikationsmodells - darin, daß sich der Lehrer als Hauptsozialisationsagent in der Schule ein möglichst genaues Bild über die Auswirkung seines Verhaltens bei den Schülern erstellen sollte. Lehrer können mit einem standardisierten, nach teststatistischen Gütekriterien erarbeitetem Beobachtungsinstrumentarium, das sie zur anonymen Beantwortung an ihre Schüler ausgeben, erfahren, wie sie von den Schülern wahrgenommen werden und entsprechende Änderungen ihres Verhaltens anstreben und durch erneute objektivierte Verhaltenseinstufung feststellen, welche Auswirkungen ihr geändertes Verhalten bei den Schülern gehabt hat.

Viele Untersuchungen der letzten Jahre zur sozialen Interaktion in der Schulklasse von TAUSCH et al. (s. PETERS in GAGE, 1970, S. 1930 f; TAUSCH und TAUSCH, 1970) belegen, welch eine Diskrepanz zwischen dem von den Lehrern für sich gewünschten und von ihnen beabsichtigten Verhalten und dem tatsächlich realisierten Verhalten besteht. Mit Hilfe des oben genannten Beobachtungsinstrumentariums wollen wir einerseits aufzeigen, ob und inwieweit der von Lehrern erwünschte von dem durch die Schüler

perzipierten Unterrichtsstil abweicht; andererseits möchten wir Aufschlüsse darüber erlangen, mit welcher Genauigkeit Lehrer Urteile über ihren von den Schülern perzipierten Unterrichtsstil abgeben können, d.h. inwieweit sie ihr eigenes Verhalten im Urteil ihrer Schüler einzuschätzen in der Lage sind. Aus der Vielfalt der Wirkungszusammenhänge, die sich in der Schüler-Lehrer-Beziehung abspielen, stellt die Genauigkeit der Wahrnehmung des Lehrers über sein Verhalten im Urteil seiner Schüler eine wichtige Variable im Hinblick auf eine Optimierung des Klassenklimas dar. So ist zu erwarten, daß Lehrer, die die Bewertung ihres Unterrichtsverhaltens durch ihre Schüler genau einschätzen und berücksichtigen können, eher in der Lage sind, potentiell sich ergebende Konflikte zu vermeiden.

Die bisherigen Ausführungen skizzieren den Problembereich des ersten Teils der vorliegenden Untersuchung, der schrittweisen Entwicklung eines Beobachtungsinstrumentariums zur Erfassung zeitstabiler Verhaltensmerkmale von Lehrern aus der Sicht der Schüler, in groben Umrissen. Dabei ist beabsichtigt, nur solche Verhaltensmerkmale von Lehrern in die Endform des Beobachtungsinstrumentariums (der Erziehungsskala zur Erfassung von Lehrerverhalten aus der Sicht der Schüler) aufzunehmen, die für das Lernklima und die Atmosphäre in Klassen von Bedeutung sind und bezüglich derer sich Lehrer unterscheiden.

Im weiteren Untersuchungsverlauf wenden wir uns der Frage zu, inwieweit Tendenzen der aktuellen Schülerbeurteilung bei Lehrern mit der Beurteilung von Lehrern durch die Schüler zusammenhängen.

Die Problematik, die sich aus der Schülerbeurteilung durch den Lehrer einerseits und der Bewertung des Unterrichtsstils des Lehrers aus der Sicht der Schüler andererseits ergibt, ist noch weitgehend unerforscht. Wie HOFER (1969) zeigen konnte, vollzieht sich die Schülerbeurteilung durch den Lehrer nach einem i m p l i z i t e n P e r s ö n l i c h k e i t s m o d e l l. - 5 -

Bekanntlich geben implizite Persönlichkeitstheorien Aufschluß darüber, daß Individuen bei der Eindrucksbildung und bei der Beurteilung anderer Menschen bestimmte Erwartungen hinsichtlich der Zusammenhänge von Eigenschaften und deren Organisation besitzen (Beispiele: "hübsch aber dumm"; "nett und hilfsbereit"; "rothaarig und leidenschaftlich" usw.). Das Problem der impliziten Persönlichkeitstheorien finden wir in der Literatur unter anderem bei CRONBACH (1955), BRUNER and TAGUIRI (1954) und HAYS (1958) abgehandelt. HOFER (1969) untersuchte, wie im speziellen Fall der Lehrer-Schüler-Beziehung die Theorie beschaffen ist, die der Lehrer von der Schülerpersönlichkeit besitzt. Er fand heraus, daß Lehrer ihre Schüler nach folgenden Eigenschaftsdimensionen beurteilen:

I. A r b e i t s v e r h a l t e n im Sinne "Disziplintiertheit vs. Undisziplintiertheit" mit den Eigenschaften konzentriert, pflichtbewußt, ordentlich, fleißig, aufmerksam, zuverlässig.

II. B e g a b u n g im Sinne "Geistige Regsamkeit" mit den Eigenschaften intelligent, begabt, einfallsreich, aufgeweckt.

III. D o m i n a n z im Sinne "Bescheidene Zurückhaltung vs. Geltungsstreben" mit den Eigenschaften geltungsbedürftig, ehrgeizig, dreist, selbstsicher, führend, unruhig.

IV, S o z i a l e A u f g e s c h l o s s e n h e i t vs. V e r s c h l o s s e n h e i t mit den Eigenschaften verschlossen, ungesellig.

V. E m p f i n d s a m k e i t vs. R o b u s t h e i t mit den Eigenschaften sensibel, kompliziert, unausgeglichen.

Dieses subjektive Ordnungsschema bzw. diese den Lehrern eigene, implizite Persönlichkeitstheorie ermöglicht es ihnen, die Schülerpersönlichkeit schnell und einfach, d.h. mit wenigen Informationen, zu erfassen. Eine schematische Beurteilung nach diesen Eigenschaftsdimensionen wird bevorzugt praktiziert von psychisch stabilen, rigiden und älteren Lehrern (vergl. HOFER 1969, S. 82 f).

Die Beurteilung bekannter Schüler durch ihre Lehrer wird weniger durch die Persönlichkeit der Schüler als vielmehr durch die implizite Persönlichkeitstheorie der Lehrer bestimmt. Diesen Befund erbrachte eine Untersuchung von BARTRAM, ROLLETT und SUHRE (1971), derzufolge neurotische Jungen[1] von Lehrern mit den Eigenschaften unbegabt, unsympathisch, pflichtvergessen, dumm, gleichgültig, desinteressiert und unkonzentriert gekennzeichnet werden, der Lehrer sieht sie als dreist und frech an. Neurotische Mädchen gelten als unbegabt und frech; Neurotizismus wird demnach bei Mädchen mit den Faktoren Intelligenz und Dominanz beurteilt. Extraversion wird bei Jungen nicht beurteilt; extravertierte Mädchen dagegen gelten als unsympathisch, unbegabt und geltungsbedürftig. Angstlichkeit[2] wird bei Jungen durch die Faktoren Arbeitsverhalten und Intelligenz beurteilt. Angstliche Jungen sind demnach in der Sicht des Lehrers unbegabt, dumm, gleichgültig, pflichtvergessen, unsympathisch, desinteressiert, unordentlich und unkonzentriert. Ängstliche Mädchen dagegen gelten als sensibel, sie zeigen nach Ansicht des Lehrers ein diszipliniertes Arbeitsverhalten. So wird Angstlichkeit bei Jungen extrem negativ beurteilt, bei Mädchen dagegen eher positiv.

---

1) Neurotizismus gemessen mit der Hamburger Neurotizismusskala von BUGGLE & BAUMGÄRTEL

2) Angstlichkeit gemessen mit dem Kinder-Angst-Test von THURNER & TEWES

Wie die Untersuchung von BARTRAM, ROLLETT und SUHRE (1971) zeigt, sind Lehrer nicht in der Lage, Persönlichkeitsmerkmale, wie sie sich in Testergebnissen manifestieren, zu erfassen. Dagegen beurteilen sie Arbeitsverhalten und Intelligenz der Schüler, verglichen mit Testergebnissen, weitgehend "richtig". Urteilsstruktur und auch Zensuren zeigen eine klare Beziehung zur Intelligenz bzw. Leistungsfähigkeit. Darüberhinaus ist Intelligenz nach Ansicht des Lehrers immer gekoppelt mit diszipliniertem Arbeitsverhalten.

Zusammenfassend läßt sich also herausstellen, daß Lehrer Persönlichkeitsmerkmale von Schülern nur im Zusammenhang mit Leistungsvariablen bewerten. Psychologische Forschungsergebnisse aber haben belegt, daß Leistungs- und Persönlichkeitsmerkmale voneinander unabhängig sind.

Schließt man angesichts der referierten Befunde zur impliziten Persönlichkeitstheorie von Lehrern einige Bemerkungen über Lehrerbildung und Lehrertraining an, so bleibt festzustellen, daß man auf eine flexiblere und differenziertere Sichtweise der Lehrer über die Schülerpersönlichkeit hinwirken sollte. Aufgrund seiner globalen Konfrontation mit Schülern ist der Lehrer besonders gefährdet, zu vereinfachen und schematische Urteile zu fällen. Lehrer sollten mit der geringen Interkorrelation zwischen Persönlichkeitsmerkmalen, der dynamischen Komplexität der kindlichen Personlichkeit und der Vielfalt der Wirkungszusammenhänge vertraut gemacht werden, die sich in der Lehrer-Schüler-Beziehung abspielen. Deshalb könnten sich eventuell aufzudeckende Zusammenhänge zwischen der aktuellen Schülerbeurteilung durch Lehrer aufgrund der oben genannten Eigenschaftsdimensionen des impliziten Persönlichkeitsmodells und relevanten Aspekten des Unterrichtsstils von Lehrern aus der Sicht der Schüler gewissermaßen als handlungsorientierende Informationen für ein Lehrertraining verwerten lassen.

Geht man davon aus, daß Lehrer hinsichtlich ihres Unterrichtsstils (im Sinne eines guten Klassen- und Lernklimas) von ihren Schülern mehr oder weniger günstig eingestuft werden, so liegt es nahe, danach zu fragen, zu welcher Art von aktueller Schülerbeurteilung Lehrer neigen, die von ihren Schülern bezüglich ihres Unterrichtsstils besonders günstig bzw. ungünstig bewertet werden. Dieser Frage wird im zweiten Teil der Hauptuntersuchung, der Validierungsaspekte umfaßt, anhand des vorhandenen empirischen Datenmaterials nachgegangen.

Als Erkundungsstudie ist außerdem vorgesehen, ob sich mit Hilfe des statistischen Verfahrens der Clusteranalyse, in die die oben genannten Eigenschaftsdimensionen der impliziten Persönlichkeitstheorie der Lehrer zur aktuellen Schülerbeurteilung als Prädiktoren eingehen, eindeutig unterschiedliche Lehrerbeurteilertypen identifizieren lassen. Im Anschluß daran soll überprüft werden, mit welcher Genauigkeit etwaige unterscheidbare Lehrerbeurteilertypen Angaben über ihren von den Schülern perzipierten Unterrichtsstil machen können, d.h. inwieweit sie ihr eigenes Verhalten in der Sicht ihrer Schüler einzuschätzen in der Lage sind.

Bevor jedoch alle speziellen Fragestellungen formuliert und durch die Auswertung des erhobenen Datenmaterials geklärt werden, referieren und diskutieren wir Literaturergebnisse zur Lehrer-Schüler-Interaktion und im weiteren Sinne zur Erziehungsstilforschung, soweit sie für unser theoretisches Konzept von Belang sind, vor allem unter methodologischen Gesichtspunkten. Dabei werden sowohl die Methoden zur Erforschung des Erziehungsstils von Lehrern als auch von Eltern abgehandelt und wo nötig, aufeinander bezogen, denn die Erforschung des Erziehungsstils von Lehrern und Eltern wirft (methodologisch) die gleichen Probleme auf.

## 2. Zur Methodologie

### 2.1. Allgemeine Bemerkungen zur Konstruktion von Beobachtungsschemata

Verhaltensmessungen im Klassenraum erfolgen gewöhnlich in zwei Schritten: Im ersten Schritt erhält man ein Protokoll zu messenden Verhaltens, im zweiten Schritt erfolgt die Quantifizierung des Protokolls. Zuvor jedoch ist zu entscheiden, welche Aspekte beobachtbaren Verhaltens protokolliert bzw. registriert werden sollen. Diese Entscheidung wird vom Ziel der Untersuchung her bestimmt.

Verglichen mit Interaktionsprozessen zwischen Schülern und solchen zwischen Lehrern sind Lehrer-Schüler-Interaktionsprozesse im Hinblick auf das Klassenklima recht gut erforscht. Man geht dabei von der deskriptiven Erfassung des Interaktionsstils der Lehrer aus und setzt ihn mit Verhaltens- bzw. Erziehungseffekten beim Schüler in Beziehung. Zweckmäßigerweise werden von den zahlreichen Verhaltensvariablen des Lehrers diejenigen ausgewählt, die bei angemessener Beobachtbarkeit, Auswertbarkeit und Zuverlässigkeit über die Zeit mit relevanten (gewöhnlich sozialen und intellektuellen) Verhaltensweisen der Schüler in enger Beziehung stehen. Hierbei kann der Untersucher, von einer Theorie ausgehend, Angaben machen, welche Verhaltensweisen im Hinblick auf das sog. Kriterium (Erziehungseffekte bei den Schülern) Lehrer am besten charakterisieren. Seine Theorie kann ihn beispielsweise dazu veranlassen, Lehreräußerungen als lehrer-, schüler- oder lernzentriert zu klassifizieren (vergl. u.a. WITHALL, 1949). Auch FLANDERS (1960) ging von einem theoretischen Bezugsrahmen aus, indem er Interaktionen als Wiederspiegelung direkten oder indirekten Lehrereinflusses einstufte.

Andererseits lassen sich in der Literatur Untersuchungen finden, die gewissermaßen "theoriefrei" Lehrerverhalten

erfassen. So erprobten JAYNE (1945) und MORSH (1956) eine Vielzahl von Items nach dem "Schrotprinzip" und hofften, für die im Schülerverhalten sich widerspiegelnde Lehreraktivität relevante Items zu ermitteln. Sie verzichteten darauf, ihren Itempool von vornherein durch eine Theorie einzuengen.

Zur Konstruktion von Items für einen Beobachtungsplan lassen sich Kategorien- und Merkmalsysteme als zwei methodische Ansätze zur Registrierung beobachtbaren Verhaltens unterscheiden (vergl. MADLEY und MITZEL in GAGE, 1965).

Ein Kategoriensystem besteht aus einer Gruppe von Kategorien, in die bestimmte Ausschnitte, Aspekte oder Verhaltensweisen eines bestimmten Typs klassifiziert werden. So wird z.B. beim BELU-V-1-D, einem Beobachtungsverfahren zur Erfassung der Lernsituation und Lernatmosphäre im Unterricht von GRUBITSCH (1970), der Kommunikationsprozeß zwischen Lehrer und Schülern in einer Unterrichtsstunde als eine Aneinanderreihung vieler einzelner, zeitlich mehr oder weniger umgrenzter sprachlicher Aktionen aufgefaßt. Dieses Beobachtungsinstrument sieht zur Erfassung unterrichtlichen Verbalverhaltens 19 Kategorien vor, die dazu verwendet werden, lehrer- und schülerbedingte "Eingänge" in verbale Interaktionsereignisse bzw. "Ausgänge" aus verbalen Interaktionsereignissen zu kennzeichnen.

Klassische Beispiele für Kategoriensysteme sind die Beobachtungspläne von WITHALL (1949), HUGHES (1959), SMITH (1959), WRIGHT and PROCTOR (1961) und FLANDERS (1960).

Merkmalssysteme sind dadurch charakterisiert, daß Verhaltensmerkmale, die während eines bestimmten Beobachtungsausschnittes auftreten, im voraus bestimmt und aufgelistet

werden. Der Beobachter hat zu registrieren, welches dieser Merkmale zur Zeit des Beobachtungsausschnittes auftritt, ggf. wie oft es auftritt. Das schon erwähnte Verfahren von MORSH (1956) ist ein Beispiel für ein Merkmalsystem.

Während bei einem Merkmalsystem viele Äußerungen bzw. Verhaltensweisen nicht berücksichtigt werden, registriert man bei einem Kategoriensystem Verhaltensweisen der zu protokollierenden Art vollständig. Im Falle der vorliegenden Untersuchung, der Erfassung zeitstabiler Verhaltensmerkmale von Lehrern aus der Sicht der Schüler, wird mit einem Merkmalssystem gearbeitet.

## 2.2. Erhebungsmethoden zur Erfassung des Erziehungsstils

### 2.2.1. Unmittelbare Beobachtung

Die ersten wichtigen Untersuchungen zum elterlichen Erziehungsstil und zur Charakterisierung von Lehrerverhalten wurden mit Hilfe systematischer, unmittelbarer Verhaltensbeobachtung durchgeführt (Life-Daten). Die im vorangegangenen Kapitel 2.1. referierten, als Merkmals- und Kategoriensysteme unterschiedenen Beobachtungspläne sind Beispiele zur Erfassung des Lehrerverhaltens durch unmittelbare Beobachtung. Der hohe Aufwand dieser Methoden führte jedoch allmählich dazu, zur Datenerhebung die ökonomischeren Fragebogen (Questionnaire-Daten) zu entwickeln, von denen in den beiden nächsten Kapiteln die Rede sein wird.

Nachteilig kann sich bei der unmittelbaren Beobachtung auswirken, daß sie für den Lehrer wissentlich erfolgt, der dadurch sein übliches Verhalten in Richtung eines erwünschten Verhaltens ändern kann. In den letzten Jahren wurde, insbesondere von TAUSCH et al. (1970), das Ausmaß komplexer Verhaltensmerkmale, wie z.B. gefühlsmäßig warm/gefühlsmäßig kalt, wertschätzend/geringschätzend,

stimulierend/nicht stimulierend, aktiv/passiv, systematisch-planend/ungeordnet-planlos usw. mit Hilfe von Schätzskalen eingestuft. Hierdurch werden Aussagen über subtile Vorgänge des Erziehungsgeschehens möglich, insbesondere dann, wenn jedes Merkmal von mindestens zwei unabhängigen Beobachtern eingeschätzt und eine bessere Übereinstimmung der Beobachter durch Vortraining ermöglicht wird.

### 2.2.2. Selbstbeschreibung

Hinsichtlich ihres Verhaltens in Klassen, ihrer Erziehungseinstellung und ihrer Persönlichkeitsmerkmale können sich Lehrer selbst beschreiben, vorausgesetzt, daß sich gewisse Antworttendenzen, wie vor allem die Neigung zur sozialen Erwünschtheit, zur Acquiescence (Ja-Tendenz) und zur Extremtendenz, kontrollieren lassen.

Unter sozialer Erwünschtheit (desirable responding) wird die Tendenz verstanden, Fragebogen-Items so zu beantworten, daß die Antwort - wie man glaubt - dem sozialen Konsens entsprechend erwünscht bzw. normenkonform ist (vergl. EDWARDS 1953 und 1957, WIGGINS 1959 und DAHLSTROM 1962).

Bei der Ja-Tendenz werden die Antwortalternativen "ja", "stimmt", "richtig" vorwiegend gewählt. Hierbei kommt die Ja-Antwort nicht durch das Wissen um die richtige Beantwortung zustande. Man sagt vielmehr gewissermaßen auf Verdacht "ja" (vergl. LENTZ 1938, CRONBACH 1942).

Extremtendenz bedeutet, daß bei Fragen vom Mehrwahltyp eher die extremen als die gemäßigten Antwortalternativen gewählt werden (vergl. CRONBACH 1946, BORGATTA und GLASS 1961, HERRMANN 1964).

Diese drei genannten klassischen Reaktionseinstellungen werden bei DAMARIN und MESSICK (1965) und RORER (1965) gründlich unter dem Gesichtspunkt diskutiert, ob sie konsistente und stabile Persönlichkeitsmerkmale sind und ob sie Rückschlüsse auf basalere Persönlichkeitsmerkmale erlauben. Die genannten Autoren stellen heraus, daß Reaktionseinstellungen faktoriell nicht einheitlich sind; beispielsweise gibt es nicht die "Ja"-Tendenz, sondern mehrere "Ja"-Tendenzen. Damit hängt zusammen, daß die empirischen Variablen, die Indikatoren für postulierte Reaktionseinstellungen sein sollen, stets bedeutsame Ladungszahlen bei mehreren Persönlichkeitsfaktoren besitzen. So soll die Ja-Tendenz einerseits mit "low verbal interpretative skill", zum anderen mit Impulsivität und temperamentsspezifischer Oberflächlichkeit zusammenhängen. Die Tendenz zur erwünschten Antwort kovariiert nach Auffassung der oben genannten Autoren hinlänglich verläßlich einerseits mit dem Mangel an wacher und kritischer Selbstbewertung und andererseits (negativ) mit einer aus Ängstlichkeit herrührenden Bewertung der eigenen Person. Die Extremtendenz soll zum einen mit einem selbstgewissen Bedürfnis nach Klarheit und Deutlichkeit, zum anderen mit einem Mangel an verbaler Nuancierungsfähigkeit korrelieren.

In diesem Zusammenhang sei eine Untersuchung neuesten Datums von NIEMANN (1972) erwähnt, die die Entwicklung von Einstellungsskalen zum Lehrer-Engagement bei Lehrerinnen in der ersten Grundschulklasse zum Gegenstand hat und die Tendenz zur sozialen Erwünschtheit als Kontrollvariable einbezieht. NIEMANN extrahierte folgende drei Faktoren:

I. Aktivität im bildungspolitischen Geschehen (Itembeispiel: "In jeder Grundschule sollten sich meiner Meinung nach die Lehrer zu Arbeitsgruppen zusammenschließen, um ein Erziehungsprogramm für sozio-kulturell benachteiligte Kinder zu entwickeln").

II. Erweiterung der beruflichen Aufgaben auf persönliche und familiäre Probleme der Kinder
(Itembeispiel: "Ich bin der Meinung, daß der Lehrer auch fürsorgliche Probleme übernehmen muß").

III. Ablehnung von stoffplanmäßig nicht festgelegten Aufgaben
(Itembeispiel: "Um private Probleme eines Kindes sollte sich ein Lehrer nur dann kümmern, wenn die ganze Klasse davon betroffen ist").

Der oben genannte Verfasser wies nach, daß erhebliche Unterschiede zwischen Lehrerinnen bezüglich ihrer Angepaßtheit und der Tendenz, sozial erwünscht zu antworten, bestehen. Durchschnittlich sind 42 % der Varianz der Reaktionen auf die Aussagen durch die Erwünschtheit der Aussagen zu erklären.

Angesichts der genannten Schwierigkeiten bei der Selbstbeschreibungsmethode ist es nach TAUSCH (1970, S. 148) "nicht verwunderlich, daß in diesen Bereichen geringe Fortschritte erzielt wurden. Andererseits ist der Zusammenhang des Lehrerverhaltens mit sog. Persönlichkeitseigenschaften von besonderem Interesse, weil hierdurch eine Möglichkeit der Voraussage des späteren Lehrerverhaltens bei Lehrerstudenten gegeben wäre."

### 2.2.3. Einschätzung durch Schüler

Eine methodologische Variante ist die Untersuchung des vom Kind wahrgenommenen Erzieherverhaltens. Eine theoretische Begründung für diese Vorgehensweise ist in sozialpsychologischen Befunden zu suchen, in denen gezeigt wird, daß in die Wahrnehmung selektive Momente eingehen, die nach SEARD und BACKMANN (1964, S. 14) von der Einstellung, Motivation und Persönlichkeit abhängen. So beeinflußt nach AUSUBEL et al. (1954, S. 173) das Erzieherverhalten

"the child's ego development only to the extent and in the form in which he perceives it", und es kann vermutet werden, daß mit der Untersuchung von perzipiertem Erziehungsverhalten der wichtigere Aspekt der Erziehungsstile gemessen wird.

Im Eltern-Kind-System zeitigt der Vergleich von perzipiertem und von Eltern berichtetem Erziehungsverhalten unbefriedigende Ergebnisse. Kinder sind ehrlicher (BROWN et al., 1947) und in ihrer Wahrnehmung genauer.

Für eine Optimalisierung der Lehrer-Schüler-Interaktion hat die Methode der Schülereinschätzungen des Lehrers befriedigende Ergebnisse erbracht. Die Rückmeldung, die der Lehrer aus anonymen Schätzurteilen seiner Schüler erhält, konnen ihn zur konstruktiven Änderung seines Unterrichtsstils veranlassen. So zeigt sich beispielsweise in einem Experiment von GAGE, RUNKEL und CHATTERJEE (1960), daß Lehrer von sechsten Schuljahren nach Erhalt von Schätzurteilen der Schüler ihr Verhalten in Richtung auf den idealen Lehrer hin änderten.

REMMERS in GAGE (1965) weist darauf hin, daß die Forschung bewiesen hat, "daß die Beurteilung durch Schüler bzw. Studenten ein zweckmäßiges, praktisches, reliables und valides Hilfsmittel für die Selbstkontrolle und die Vervollkommnung der Lehrer ist" (REMMERS in GAGE, 1965, S.962). Folgende wichtigste allgemeine Ergebnisse dieser Forschung lassen sich herausstellen (zitiert nach REMMERS in GAGE, 1970, S. 963):

- "Die Reliabilität der Schätzurteile von Studenten über Lehrer ist im Sinne der Spearman-Brown-Formel eine Funktion der Anzahl der Beurteiler. Wenn man 25 oder mehr Studentenurteile mittelt, sind sie genau so

reliabel wie die gegenwärtig verfügbaren besseren Schulleistungs- und Intelligenztests."

- "Wenn überhaupt, haben die Leistungsnoten der Studenten nur wenig Einfluß auf die Beurteilung der Lehrkräfte, die die Noten erteilt haben."

- "Ehemalige Studenten stimmen zehn Jahre nach Abgang beachtlich mit der Beurteilung derselben Lehrkräfte durch noch im Studium befindliche Studenten überein."

-"In ein und derselben Hochschule können hinsichtlich der Effektivität der Veranstaltungen im Urteil der Studenten zahlreiche und erhebliche Unterschiede zwischen den Instituten bestehen."

- "Das Geschlecht der studentischen Beurteiler hat keinen oder kaum einen Einfluß auf die Einschätzung der Lehrer."

- "Das Geschlecht der Lehrkräfte spielt für die Beurteilung im allgemeinen keine Rolle."

-"Der für eine solche Beurteilung nötige Aufwand ist gering. Er ist sogar beträchtlich niedriger als die Kosten, die mit der Durchführung eines umfangreichen standardisierten Leistungstest der üblichen Art verbunden sind."

Weitere Hinweise auf Untersuchungen zur Einschätzung des Lehrerverhaltens durch Schüler sind REMMERS (a.a.O.) und TAUSCH (1970, S. 146f.) zu entnehmen.

### 2.2.4. Kritische Würdigung der Erhebungsmethoden

Die von den vorliegenden Erhebungsmethoden gelieferten unterschiedlichen Daten lassen sich nach HERRMANN (1969, S. 164-190) wie folgt einteilen:

- Life-Daten (Fremdbeurteilung): Verhaltensbeobachtung
- Questionnaire-Daten (Selbstbeurteilung): Fragebogen, Retrospektion, Interview.

wegen der Abhängigkeit der Fremd- und Selbstbeurteilungen von dritten Variablen wie z.B. von Beantwortungsstilen und Beurteilungsfehlern führen sie selten und nur unter bestimmten Voraussetzungen zu übereinstimmenden Ergebnissen. Erstrebenswert, wenn in der Erziehungsstilforschung kaum verwendet, sind T-Daten (Testdaten), denen der höchste Grad der Objektivität zukommt und die zu den Q-Daten in einem vermuteten Kovariationszusammenhang stehen: Faktoren zweiter Ordnung aus Q-Daten korrelieren mit Faktoren erster Ordnung aus T-Daten (vergl. HERMANN 1969, S. 189).

## 2.3. Das Konzept vom Erziehungsstil

Erziehungs- oder Interaktionsstile sind auffaßbar als statistische Klassen erziehungsrelevanter Verhaltensweisen und konnen in Erziehungseinstellungen (Vorstellungen und Normen) und Erziehungspraktiken untergliedert werden. Eine haufig zitierte Definition von Erziehungsstil ist bei EYFERTH (1966, S. 23) zu finden: "Als Erziehungsstil wird eine Gruppe von Merkmalen des Erziehungsverhaltens bezeichnet, in welcher großere gemeinsame Merkmalsvarianz herrscht, als nach Variabilitat aller Merkmale zustande kommen könnte, und welche Gruppe die Eigenarten in diesem Stile Erzogener genauer vorauszusagen erlaubt als Einzelmerkmale." In dieser Definition wird also die überzufällige Korrelation einzelner Erziehungsmerkmale gefordert, von denen man einen Präzisionsgewinn bei der Vorhersage der Erziehungseffekte erwartet.

Um was für Konglomerate von Einzelmerkmalen, die hier als Erziehungsstile angesprochen werden, handelt es sich?

In der einschlägigen Literatur werden der autokratische und sozialintegrative (oder demokratische) Stil als zwei Hauptvarianten erzieherischen Verhaltens herausgestellt. Beide Stile - was deren Einzelmerkmale anbelangt - sind insbesondere durch das Sprachverhalten von Lehrern Schülern gegenüber gekennzeichnet. Das Merkmal der Höflichkeit gilt beispielsweise als konstitutiv für den sozialintegrativen Stil und ist u.a. durch den häufigen Gebrauch der Worte "bitte" und "danke" gekennzeichnet. Insgesamt zeichnet sich der sozialintegrative Erziehungsstil dadurch aus, daß der Lehrer einen Überblick über Ziele und Vorhaben gibt, die dann Gegenstand von Gruppendiskussionen und Entscheidungen werden. Der Lehrer übernimmt die Entscheidungen nur dort, wo es unumgänglich erscheint. Es wird im allgemeinen großes Verständnis gegenüber den Wünschen der Schüler gezeigt. Der autokratische Erziehungsstil dagegen beinhaltet intensive Lenkung durch den Erzieher mit deutlicher Dominanz des Erziehers. Man findet häufiger Befehle, Drohungen, Strafen, die Aktivität und Spontaneität der Schüler tritt in den Hintergrund, der Lehrer bestimmt im wesentlichen, und der Unterricht besteht größtenteils aus Reden und Fragen des Lehrers.

Es stellte sich heraus, daß die Art der Erhebungsmethode, mit der der elterliche - und Lehrererziehungsstil erfaßt werden sollte, nicht ohne Einfluß auf die Untersuchungsergebnisse geblieben ist. Nach EYFERTH (a.a.O., S. 25 f) lassen sich die Erhebungsinstrumente in zwei Gruppen einteilen:

- Verhaltenseinstufungen bzw. Beobachtungen des wahrnehmbaren Erziehungsverhaltens;
- Fragebogen und Interview.

Die Entwicklung der FELS - PARENT - BEHAVIOR- SCALES (1941), des Fragebogens von SHOBEN (1949) und des PARI (Parental Attitude Research Instrument) von SCHAEFER und BELL (1958) bis hin zu BRONFENBRENNERS PARENTAL BEHAVIOR QUESTIONNAIRE (s. SIEGELMAN, 1965) zeigt im Bereich der Erforschung des elterlichen Erziehungsstils deutlich das Bemühen, einerseits zu einwandfreien Testinstrumenten zu kommen, andererseits inhaltlich detaillierter die in Frage stehenden Merkmale des Erziehungsgeschehens anzugehen.

Mit den genannten Instrumentarien hat man aufgrund von empirischen Untersuchungen Datensammlungen angelegt, die ex post interpretiert und Ad-hoc-Modellen zugeordnet wurden, wenn man überhaupt über die reine Datenbeschreibung hinausging. Ein solches Vorgehen birgt in sich bestimmte Schwächen: Sobald die mit einer bestimmten Population und aufgrund einer bestimmten Erhebungsmethode zustandegekommenen Untersuchungsergebnisse generalisiert werden, zeigen sich - im Vergleich zu anderen Autoren - zunächst rein sprachliche Schwierigkeiten, vor allem aber widersprüchliche Ergebnisse, die auf die Unwägbarkeiten der untersuchten Stichprobe, des Instrumentes und auf das oft undurchsichtige Rahmenkonzept für die Erklärung und Interpretation zurückzuführen sind. So ist es angesichts faktorenanalytischer Untersuchungen wichtig, darauf hinzuweisen, daß unterschiedliche Faktorenstrukturen, d.h. unterschiedliche Anzahl relevanter Faktoren, unterschiedliche Interpretation der Faktoren und unterschiedliche Beiträge aufgeklärter Varianz bei Faktorenanalysen davon abhängen, auf welche Klasse empirischer Daten und auf welche Referenzpopulation man sich bezieht. Deshalb ist es wichtig, sich zu vergewissern, ob man als Referenzpopulation Eltern, Lehrer, Kinder oder unabhängige Beobachter hat, als Klasse empirischer Daten Erziehungshaltungen oder -praktiken.

Obwohl deshalb in den Erziehungswissenschaften noch keine einheitliche, allgemeinverbindliche Taxonomie der Erziehungsvorstellungen und des Erziehungsverhaltens vorhanden ist, sind doch aufgrund von dimensionalanalytischen Untersuchungen bestimmte basale Faktoren häufiger repliziert worden, wie vor allem "Autonomie vs. Kontrolle" und "Zuwendung vs. Zurückweisung" (vergl. HERRMANN, 1969, S. 348 f.). BECKERS (1964) bekanntgewordenes Modell zur Beschreibung elterlichen Verhaltens unterstellt die drei orthogonalen Faktoren "Förderung vs. Hemmung der kindlichen Selbstverwirklichung", "emotionale Wärme vs. Feindseligkeit" und "gelassene Abständigkeit vs. ängstliches Verwickeltsein". SIEGELMANN (1965, 1966) beschreibt vom Kinde erlebtes elterliches Erziehungsverhalten durch die drei Faktoren "liebevolle Zuwendung (loving)", "Strenge (punishment)" und "elterliche Forderungen (demanding)".

Wesentliche Merkmale des durch sprachliche Äußerungen dokumentierten Lehrerverhaltens (z.B. Häufigkeit der Lehrerfragen, -befehle und -aufforderungen und Anteil der irreversiblen Äußerungen) kehren weitgehend unabhängig von Unterrichtsinhalt, Lerngruppe und Erhebungsinstrumenten bei denselben Lehrern in weitgehender Konstanz wieder. Faßt man die Ergebnisse vieler Arbeiten vereinfacht zusammen (vergl. TAUSCH und TAUSCH, 1970), so ergeben sich (faktorenanalytisch) deutlich zwei Hauptdimensionen des Lehrer - Erzieherverhaltens:

1. Die emotionale Dimension: Wertschätzung (emotionale Wärme, Zuneigung) vs. Geringschätzung (emotionale Kälte und Abneigung).

2. Die Lenkungsdimension: Maximal starke Lenkung (Kontrolle, Restriktion) vs. minimale Lenkung (etwa Permissivität, Autonomie, Gewährenlassen, minimale Kontrolle).

Im übrigen können zusätzlich noch die folgenden Dimensionen unterschieden werden: "Aktivität vs. Passivität", zum Teil übereinstimmend mit der Dimension "Engagement vs. Desinteressiertheit". Für die spezifisch unterrichtliche Tätigkeit ergibt sich schließlich noch die Dimension: "Ausgeprägte Klarheit und Ausdrucksfähigkeit vs. Unklarheit und Verschwommenheit im unterrichtlichen Vorgehen".

Vereinfacht lassen sich die Auswirkungen des dimensionsanalytisch beschriebenen Lehrerverhaltens auf das Schülerverhalten wie folgt kennzeichnen:
Bei hohem Ausmaß an emotionaler Wärme, Wertschätzung und Zuneigung von Lehrern entstehen bei Schülern mit hoher Wahrscheinlichkeit:

a) Positive reziproke Affekte (emotionale Sicherheit, Gefühl der Wertschätzung, Möglichkeiten zur angstfreien Erziehung);
b) Bekräftigung des Selbstwerterlebens;
c) Geringe Aggressivität zwischen den Schülern.

Bei hohem Ausmaß an Lenkung, Dirigismus und Kontrolle ist auf Seiten der Schüler zu erwarten:

a) Geringes Ausmaß an spontanem und kreativem Verhalten;
b) Größeres Ausmaß an Spannungen;
c) Geringes Gefühl von Freiheit;
d) Höheres Ausmaß an Leistung in bestimmten Bereichen;
e) Übernahme dirigistischer Führungsstile.

Bei einem mittleren bis geringen Ausmaß an Lenkung, Dirigismus und Kontrolle ergeben sich die entgegengesetzten Effekte.

## 2.4. Lernpsychologische Spezifizierung eines Erziehungsstilmodells als theoretischer Ausgangspunkt der Untersuchung

Die Ausführungen des vorangegangenen Kapitels 2.3. sollten verdeutlichen, daß die operationalen Definitionen von Erziehungs- und Interaktionsstilen die Problematik der begrifflichen und empirischen Abgrenzungen erziehungsrelevanter Verhaltensklassen implizieren und daß unter den üblichen Forschungsansätzen zur Aufklärung der Strukturiertheit des Erziehungsstils die deskriptiven Klassifikationen mit Hilfe statistischer, meist faktorenanalytischer Verfahren dominieren. Bei diesen deskriptiv-klassifikatorischen (faktorenanalytischen) Strategien werden a posteriori Zusammenhänge aus vorliegenden Daten festgestellt und dann mit Erziehungseffekten in Beziehung gesetzt.

Im Sinne einer wissenschaftstheoretisch einwandfreien Grundlegung eines Erziehungsstilmodells ist - angesichts der bisher referierten Ex post-Modelle - eher ein A priori-Modell angezeigt, das unter anderem durch lerntheoretische Konzepte fundiert sein kann, aus denen sich die spezifischen Effekte bei den Erzogenen a b l e i t e n lassen. Eine psychologische Lerntheorie genügt deshalb dieser Anforderung, weil sie sich aus einigen basalen Prinzipien der Allgemeinen Psychologie herleiten läßt.

SECORD und BACKMAN (1964, S. 527 f) nennen die folgenden für den Sozialisationsbereich relevanten Lernprinzipien: Operantes Lernen, intentionales und inzidentelles Lernen, Instruktionslernen, Modellernen. Diese unterschiedlichen Elementarformen des Lernens haben sich in ein übergeordnetes Konzept bisher nicht integrieren lassen. Vorzugsweise wurde das operante Lernen zu Modellbildungszwecken aufgegriffen. So hat, ausgehend von der Zwei-Faktorentheorie MOWRERS (1960) HERRMANN (1968, 1971) ein zweidimensionales Modell elterlicher Bekräftigungsweisen konzipiert, das die orthogonalen Dimensionen elterlicher Strenge (Strafe, Druck, Tadel) und elterliche Unterstützung (Lob, Hilfe,

Förderung) unterstellt. Zur operationalen Erfassung entwickelten HERRMANN et al. (1971)entsprechend eine Strenge- und Unterstützungsskala, die sich an das "Parent Behavioral Questionnaire" (BPB) anlehnt. Unter Strenge bzw. Unterstützung (positiver bzw. negativer Bekräftigung) ist - auf die konkrete Erziehungspraxis bezogen - die Unterstützung von erwünschten bzw. die Unterbindung von unerwünschten kindlichen Verhaltensweisen mit verbalen und instrumentellen Mitteln zu verstehen.

Nach lerntheoretischen Prinzipien bewirken positive und negative Bekräftigung unterschiedliche Effekte beim Lernenden bzw. Erzogenen, wobei positive Bekräftigung (Belohnung) zur Vergrößerung der Auftretenswahrscheinlichkeit belohnten Verhaltens und zum Aufsuchen der Belohnungsquelle und negative Bekräftigung zur Verminderung der Auftretenswahrscheinlichkeit des bestraften Verhaltens und zur Meidung der Bestrafungsquelle führt.

Das soziale Verhalten vorwiegend unterstützend erzogener (positiv bekräftigter) Individuen einerseits und vorwiegend streng erzogener (negativ bekräftigter) Individuen andererseits läßt sich mit Hilfe basaler lerntheoretischer Überlegungen genauer spezifizieren. So führt, wenn wir den Ausführungen STAPFS et al. (1971) folgen, vorwiegend unterstützendes Erziehungsverhalten beim Erzogenen zur G e b o t s o r i e n t i e r u n g, zur Neigung, sozial erwünschte Normen der Bezugsgruppe auszuführen. Für unterstützend und vorwiegend belohnend erzogene Individuen ist es charakteristisch, das sozial erwünschte Verhaltensweisen durch differentielle positive Bekräftigungen, d.h. sukzessive Verhaltensausformung ( S h a p i n g) spezialisiert und damit "optimalisiert" werden. Gebotsorientierte, die während ihrer Lerngeschichte relativ viele positive Bekräftigungen erhielten, haben viele differenzierte und ausgeformte Fertigkeiten, die in ihrer

Bezugsgruppe als "richtig" angesehen werden und zu denen unter anderem eine elaborierte Sprache und das Beherrschen sozialer Spielregeln und Rituale (gutes Benehmen usw.) gehören. Die relativ vielen positiven Bekräftigungen, die dem Gebotsorientierten im Laufe seiner Lerngeschichte zuteil werden, führen zu seiner hohen generellen Aktivität und seiner hohen Frustrationstoleranz. Das soziale Verhalten des Gebotsorientierten ist auch beim Übergang in eine neue soziale Bezugsgruppe ausgeprägt aktiv aufsuchend, unbefangen und prononciert.

Vorwiegend streng erzogene (negativ bekräftigte) Individuen dagegen neigen zur V e r b o t s o r i e n t i e r t h e i t, sie meiden Reaktionen, die nach den Normen der Bezugsgruppe unerwünscht sind. Häufige negative Bekräftigung während ihrer Lerngeschichte führt bei ihnen zu Vermeidungsverhalten und zu einer generellen Aktivitätsminderung. Der soziale Verhaltensstil des Verbotsorientierten ist vorwiegend zurückhaltend und abwartend, seine Verhaltensalternativen sind, als Folge vielfältiger und häufiger Bestrafung, gehemmt, wenngleich nicht "endgültig" gelöscht. Da die Verbotsorientierten häufige Bestrafung erfahren haben, werden von ihnen negative Verhaltenskonsequenzen antizipiert, die sie aktiv vermeiden, wobei die Vermeidungstendenz unter anderem im Verzicht auf Abweichen und auf Risiko besteht; hinsichtlich ihrer sozialen Verhaltensweisen halten sie sich betont im "Feld des Erlaubten" auf, wobei "Grenzüberschreitungen" unter anderem durch strikte Übernahme von Autoritäts- und Mehrheitsmeinungen vermieden werden. Beim Eintritt in neue soziale Bezugsgruppen ist die Frustrationstoleranz beim Verbotsorientierten nicht groß, wenn Frustration die plötzliche Bestrafung zuvor bekräftigter Verhaltensweisen bedeutet, denn der Verbotsorientierte ist, da bei ihm vorherige negative Bekräftigungen überwiegen, durch schwache soziale Aktivität charakterisiert, so daß von

daher schon geringe Wahrscheinlichkeit besteht, in der "neuen" Gruppe Unerwünschtes zu tun. Wenn aber andererseits der Verbotsorientierte wahrnimmt, daß bislang bestrafte Verhaltensweisen in einer "neuen" Bezugsgruppe erwünscht sind und belohnt werden, dann können zuvor negativ bekräftigte (unterdrückte) Verhaltensalternativen undifferenziert und ungehemmt "ausbrechen". Lernpsychologen sprechen dann von einer "sekundären" (undifferenzierten) Aktivitätsvermehrung, da zuvor bestrafte, unerwünschte Verhaltensalternativen nicht durch sukzessive Verhaltensausformung (S h a p i n g ) ausdifferenziert worden waren (vergl. u.a. ESTES, 1964 und 1969, AZRIN und HOLZ, 1966).

Die lernpsychologisch ableitbaren Verhaltensmerkmale des Verbotsorientierten lassen sich zusammenfassend als " B r a v h e i t s s y n d r o m" kennzeichnen, wozu die primäre Aktivitätsminderung, der Verzicht auf Abweichung und Risiko und die sekundäre Aktivitätssteigerung (beim Übergang in "neue" soziale Bezugsgruppen) gehören. Andererseits lassen sich die hohe (primär erworbene) Aktivitätsrate und das aufsuchende Verhalten beim Gebotsorientierten, seine spielerische Beherrschung kultureller Techniken, sozialer Regeln und Rituale und die starke Verhaltensdifferenzierung durch Shaping, aber auch seine erhöhte Frustrationstoleranz beim Normenwechsel in "neuen" sozialen Bezugsgruppen unter dem sog. "C l e v e r n e s s s y n d r o m" subsumieren (vergl. STAPF et al., 1971, S. 40 - 46).

## 2.5. Operationalisierung von Strenge und Unterstützung

In diesem Abschnitt soll das bisher lernpsychologisch begründete Erziehungsstilkonzept positiver und negativer Bekräftigungsweisen methodologisch weiter spezifiziert und die operationalen Definitionen des theoretischen Begriffsinventars dargelegt werden.

Wir hatten negative und positive Bekräftigungen als relevante Merkmale des Erziehungsverhaltens herausgestellt. Sie werden vom Kinde als Strenge und Unterstützung erlebt. Hierbei sind Strenge und Unterstützung das Maß für die relative Häufigkeit von Bestrafungen und Belohnungen verschiedenen Intensitätsgrades.

Bei der Auswahl der Items lehnen wir uns an die von HERRMANN et al. (1971) erprobte M a r b u r g e r E r z i e h u n g s s k a l a zur Erfassung des elterlichen Erziehungsstils aus der Sicht des Kindes an und überprüfen die Brauchbarkeit dieses zweidimensionalen Konzepts Strenge und Unterstützung zur Charakterisierung des Lehrerverhaltens aus der Sicht der Schüler anhand verschiedener, noch näher zu beschreibender Schülerpopulationen.

Bei dem von uns verwendeten Instrumentarium handelt es sich um Fragebogen-Skalen. Mit ihnen registrieren wir keine Dokumentaraufnahmen des Lehrer-Erzieherverhaltens, sondern "t y p i s c h e s" Erzieherverhalten, t y p i c a l b e h a v i o r im Sinne CRONBACHS (1964); d.h.es geht dabei um das Erfassen der Wahrscheinlichkeit, mit der Lehrer als Erziehungspersonen in beliebigen Situationen zu einem bestimmten Erziehungsverhalten neigen. Angesichts des Erfassens des typischen Lehrer-Erzieherverhaltens handelt es sich nicht um die spezifische Wirkung eines ganz bestimmten Lehrerverhaltens auf bestimmte Schüler in einer bestimmten Situation, sondern um Lehrereigenschaften, die g l o b a l auf die Schüler wirken. Diese Lehrereigenschaften lassen sich am besten mit dem t r a i t -Begriff CATTELLS (1965) charakterisieren. Hierbei nehmen wir in Kauf, daß die Ausprägungsgrade von Häufigkeit und Intensität des Lehrer-Erzieherverhaltens als separate Aspekte unberücksichtigt bleiben.

Mit der Erfassung global auf die Schüler einwirkender Lehrereigenschaften geben wir jedoch das lernpsychologisch

orientierte zweidimensionale Bekräftigungskonzept nicht auf. In der psychologischen Literatur wird hinsichtlich des Bekräftigungsbegriffs zwischen k o n t i n g e n t e n und n i c h t - k o n t i n g e n t e n Verstärkungen unterschieden (vergl. u.a. ATKINSON und ROBINSON, 1961, SILVERMAN, 1963 und BLÖSCHL, 1969). Hierbei wird unter kontingent die Verstärkung einer bestimmten, relativ isolierbaren (partiellen) Reaktion im Lernexperiment verstanden. Im sozialen Kontext und in vielen empirischen Untersuchungen haben "Belohnung" und "Bestrafung" jedoch nicht die Funktion der Verstärkung spezifischer (isolierter) Reaktionen, sondern werden als Konsequenzen sehr globaler Verhaltensweisen aufgefaßt. Werden beispielsweise Schüler vom Lehrer bei Leistungsfortschritten (verbal) gelobt, oder bestraft der Lehrer bei Unruhe die ganze Klasse, so kann ebenfalls von positiven bzw. negativen Verstärkungen gesprochen werden, die allerdings nicht auf spezifische Reaktionen wie im Lernexperiment, sondern auf globale Verhaltensmuster folgen und daher nicht-kontingent sind.

Diese Hinweise mögen verdeutlichen, daß wir uns bei der Erfassung der Bekräftigungswirkung von Lehrern auf einer Verhaltensebene bewegen, die lernpsychologisch als "molar" charakterisiert werden kann.

## 2.6. Hinweise zum gesamten Untersuchungsablauf

In den folgenden Kapiteln werden die einzelnen Stufen zur Konstruktion und Erprobung der Erziehungsskala zur Erfassung zeitstabiler Verhaltensmerkmale von Lehrern aus der Sicht der Schüler abgehandelt. Dabei erfolgt die Auswahl von Items für die "T e s t e n d f o r m" (die endgültige Erziehungsskala) durch eine Vor- und Hauptuntersuchung.

In der Voruntersuchung soll die vorläufige Brauchbarkeit eines noch näher zu beschreibenden Itempools an einer

relativ altersheterogenen Stichprobe ermittelt werden. Der erste Teil der Hauptuntersuchung, in die unter anderem auch Lernbehindertenklassen einbezogen werden, soll die für alle Schultypen geeignete "T e s t e n d f o r m" erbringen.

Bei der Itemselektion zur "T e s t e n d f o r m" werden verschiedene Selektionskriterien zur Anwendung kommen, unter anderem werden insbesondere solche Items ausgeschieden, die, anstelle von Lehrerverhalten aus der Sicht der Schüler, sozial erwünschtes Antwortverhalten der Schüler erfassen. Zudem werden alle Items bezüglich ihrer Validität an Expertenurteilen überprüft (Expertenrating), wobei die für Klassenklima und Lernatmosphäre nicht validen Items ausgeschieden werden.

Vor- und erster Hauptuntersuchungsteil zur Itemselektion stellen den ersten Untersuchungsabschnitt der vorliegenden Arbeit dar. Im zweiten Hauptuntersuchungsteil wird versucht, mit Hilfe der Clusteranalyse Beurteilertypen von Lehrern zur aktuellen Schülerbeurteilung zu finden. Zur Klassifikation von Lehrern werden die fünf Beurteilungsdimensionen zur impliziten Persönlichkeitstheorie der Lehrer (nach HOFER, 1969) als Prädiktoren herangezogen. Anschließend werden bestimmte Lehrerbeurteilertypen daraufhin untersucht, wie sie einerseits von ihren Schülern bezüglich ihres Unterrichtsstils eingeschätzt werden und wie sie sich andererseits selbst im Urteil ihrer Schüler einstufen.

## II. VORUNTERSUCHUNG

### 3. Fragestellungen zur Voruntersuchung

1) Welche Aspekte des Unterrichtsverhaltens von Lehrern werden von Schülern als zusammengehörig wahrgenommen?

2) Welche Items eines zu analysierenden Itempools erweisen sich - nach aufgabenanalytischen Kriterien - für einen Schülerfragebogen als brauchbar?

3) Unterscheiden sich die von Lehrern erwünschten und geschätzten Schülerurteile von tatsächlich erhaltenen Schülerurteilen?

4) Lassen sich bezüglich erwünschter, geschätzter und tatsächlich erhaltener Schülerurteile bei Lehrern schultypen- und/oder geschlechtsspezifische Unterschiede nachweisen?

### 4. Durchführung der Voruntersuchung

#### 4.1. Zusammenstellung des zu analysierenden Itempools

Bei der Auswahl und Zusammenstellung von Items, die erlebte Strenge und Unterstützung von Lehrern messen sollen, bot sich ein bereits von BAUMANN (1971) verwendeter Itemfragebogen an. Dieser Fragebogen soll - im Sinne der oben abgehandelten, auf lerntheoretischen Erwägungen beruhenden Konzepte positiver und negativer Bekräftigung - angebrachtes bzw. unangebrachtes Interaktionsverhalten von Lehrern erfassen. Aus diesem Fragebogen entnehmen wir zehn Items, die nach der Relevanz für Strenge und Unterstützung am besten geeignet erschienen und die - als Kriterium zur Unterscheidung von Lehrern - innerhalb einer Klasse geringe und zwischen Klassen große Streuungen aufweisen. Alle weiteren Items unseres Itempools wurden in Diskussionen mit Lehrern konzipiert.

Bei jeder Einzelfrage stehen den Schülern vier Antwortalternativen zur Wahl, die als Skalenintervalle behandelt werden. Jedes Item hat, wie z.B. anhand von Item 1 deutlich wird, vier Antwortalternativen:

1. Der Lehrer freut sich, wenn er die Klasse betritt
   oft  manchmal  selten  nie

Folgende 44 Items wurden 780 Schülern und Schülerinnen zur Beurteilung ihrer jeweiligen Klassenlehrer vorgelegt:

1. Der Lehrer freut sich, wenn er die Klasse betritt.
2. Er achtet auf Pünktlichkeit.
3. Der Lehrer läßt Vieles durchgehen.
4. Der Lehrer unterstützt Schüler, die in der Klasse nicht so beliebt sind.
5. Er hat es gern, wenn wir still auf unseren Plätzen sitzen.
6. Wenn er etwas verspricht, dann hält er das auch.
7. Er gibt Klassenarbeiten schnell zurück.
8. Der Lehrer hat schlechte Laune.
9. Er achtet auf genaue Einhaltung von Pausenzeichen.
10. Der Lehrer bespricht mit uns Schülern die Auswahl der Unterrichtsthemen.
11. Der Lehrer ist auch zu den Schülern nett, die faul sind.
12. Wenn der Lehrer etwas erklärt hat, dann sagt er, daß es wohl nun auch dem Letzten klar geworden sein müsse.
13. Wenn er etwas anordnet, dann begründet er das auch.
14. Der Lehrer sagt, daß er Fragen von Schülern für dumm hält.
15. Der Lehrer erklärt schwierige Dinge und Aufgaben so, daß alle, auch die schwächeren Schüler der Klasse, es verstehen.
16. Der Unterricht ist langweilig, daß man einschläft.
17. Der Lehrer sagt uns gegenüber manchmal Dinge, die wir zu ihm nicht sagen dürfen.

18. Der Lehrer lobt uns, wenn wir gute Arbeit leisten.
19. Wenn Schüler untereinander Streit haben, dann sagt der Lehrer, wie sie sich verhalten sollen.
20. Der Lehrer stellt einige Schüler vor der Klasse bloß.
21. Auch wenn der Lehrer bei Schülern Fehler entdeckt, bleibt er freundlich.
22. Wenn wir etwas getan haben, was der Lehrer falsch findet, dürfen wir sagen, warum wir es getan haben.
23. Wenn man viele Fragen hat, dann beantwortet sie der Lehrer alle.
24. Der Lehrer erteilt Befehle.
25. Auch wenn man mal frech gewesen ist, verliert der Lehrer nicht die Ruhe.
26. Wenn wir nicht sofort tun, was der Lehrer sagt, wird er ärgerlich.
27. Der Lehrer sorgt dafür, daß alle Schüler arbeiten, so gut sie können.
28. Der Lehrer schimpft mit uns.
29. Der Lehrer ist freundlich und man hat leicht Zugang zu ihm.
30. Der Lehrer zieht einige Schüler besonders vor.
31. Der Lehrer erklärt, wie er unsere Klassenarbeiten beurteilt.
32. Der Lehrer gibt uns Aufgaben, ohne zu sagen, wie wir sie ausführen sollen.
33. Ich gehe gern zur Schule.
34. Wenn man ein Schulbuch vergißt, dann schimpft der Lehrer.
35. Der Lehrer bringt uns Dinge bei, die wir gern können möchten.
36. Der Lehrer weiß normalerweise, was er will.
37. Wenn in der Klasse Unruhe ist, dann bestraft der Lehrer die ganze Klasse.
38. Man kann mit ihm über alles sprechen, auch wenn man etwas ausgefressen hat.
39. Wenn der Lehrer bestraft, dann merkt man, daß er es nicht gern tut.

40. Der Lehrer sieht ein, daß die Schüler recht haben und gibt nach.
41. Der Lehrer gibt Strafarbeiten auf.
42. Wenn man zu spät zur Schule kommt, dann hat der Lehrer dafür Verständnis.
43. Bei unserem Lehrer können wir uns viel herausnehmen.
44. Bei unserem Lehrer dürfen wir offen unsere Meinung sagen.

## 4.2. Ablauf der Datenerhebung

Die Befragung wurde in vier Haupt- und vier Gesamtschulen im Ruhrgebiet und Sauerland durchgeführt. In Tabelle 1 sind die an der Untersuchung beteiligten Jahrgänge aufgeführt.

Tabelle 1
Zusammensetzung der Stichprobe

| | Klasse (Jahrgang) | | | | | | |
|---|---|---|---|---|---|---|---|
| Schulart | 5 | 6 | 7 | 8 | 9 | 10 | |
| Hauptschule | 1 | 2 | 2 | 4 | 2 | 1 | Σ 12 |
| Gesamtschule | 2 | 8 | 4 | - | - | - | Σ 14 |
| Σ | 3 | 10 | 6 | 4 | 2 | 1 | Σ 26 |

Um die Verwendbarkeit des Fragebogens an möglichst vielen Altersstufen zu überprüfen, wurde eine altersheterogene Stichprobe gewählt. Die VI, Kollegen der jeweilig zu beurteilenden Klassenlehrer, berichteten, daß bei der Beantwortung der Items auch in den jüngeren Klassen kaum Verständnisschwierigkeiten auftraten. Die Durchführungsdauer schwankte zwischen 15 und 30 Minuten.

Die VI wurden gebeten, solche Kollegen von ihren Schülern beurteilen zu lassen, die nach dem Eindruck der VI sehr unterschiedliche Unterrichtsstile zeigen. Durch diese Selektion sollte eine Varianzvergrößerung der Schülerurteile über die Lehrer bewirkt werden.

Im Anschluß an die Schülerbefragung wurden jedem beurteilten Klassenlehrer zwei Exemplare des Schülerfragebogens mit der Bitte vorgelegt, itemweise die durchschnittlich von ihren Schülern erhaltenen Beurteilungen zu schätzen (vermutetes Realurteil) und itemweise anzugeben, welche Idealbeurteilungen sie sich von ihren Schülern wünschen (Wunschurteil).

## 4.3. Auswertung der Daten und Darstellung der Ergebnisse

### 4.3.1. Faktorenanalyse der Items

Es wurde eine Hauptkomponentenanalyse (R-Analyse) für die Gesamtstichprobe (N = 780) durchgeführt mit dem Ziel, die Items zu Subskalen zusammenzustellen[1]. Drei Faktoren, die extrahiert und nach dem Varimaxkriterium auf Einfachstruktur rotiert wurden, ließen sich wie folgt interpretieren:

I. UNTERSTÜTZUNG (Lehrer wird als hilfreich und zielstrebig wahrgenommen)

II. STRENGE (Lehrer wird als egozentrisch-dirigistisch wahrgenommen)

III. AUSGLEICHENDES GEWÄHRENLASSEN (Lehrer wird als nachgiebig-zuwendend wahrgenommen)

Wie aus Tabelle 2 ersichtlich wird, weisen einige Items auf zwei Faktoren etwa gleich hohe Ladungszahlen auf. Diese Items ordneten wir bei der Subskalenzuweisung jeweils der Dimension zu, auf der sie am höchsten laden.

1) Alle erforderlichen Rechenarbeiten, für die wir Herrn Manfred Tücke zu Dank verpflichtet sind, wurden mit Bibliotheksprogrammen des Psychologischen Instituts der Universität Marburg an der Zentralen Rechenanlage in Marburg durchgeführt.

Tabelle 2

Schwierigkeiten, Trennschärfenindizes und Faktorenladungen der Items

| I. Unterstützung | P | $r_{it}$ | Faktor I | Faktor II | Faktor III |
|---|---|---|---|---|---|
| 1. Der Lehrer freut sich, wenn er die Klasse betritt. | 65 | .49 | .33 | -.22 | .26 |
| 2. Wenn es etwas verspricht, dann hält er das auch. | 89 | .44 | .45 | -.13 | .07 |
| 3. Er gibt Klassenarbeiten schnell zurück. | 67 | .37 | .31 | -.13 | -.06 |
| 4. Wenn er etwas anordnet, dann begründet er das auch. | 90 | .39 | .34 | -.03 | .14 |
| 5. Der Lehrer erklärt schwierige Dinge und Aufgaben so, daß alle, auch die schwächeren Schüler der Klasse, es verstehen. | 91 | .36 | .35 | -.07 | .10 |
| 6. Der Unterricht ist langweilig, daß man einschläft. | 63 | .50 | -.43 | .19 | -.07 |
| 7. Der Lehrer sagt uns gegenüber manchmal Dinge, die wir zu ihm nicht sagen dürfen. | 76 | .45 | -.43 | .38 | -.02 |
| 8. Auch wenn der Lehrer bei Schülern Fehler entdeckt, bleibt er freundlich. | 87 | .57 | .34 | -.29 | .30 |
| 9. Der Lehrer ist freundlich, und man hat leicht Zugang zu ihm. | 27 | -.15 | .47 | -.22 | .38 |
| 10. Der Lehrer zieht einige Schüler besonders vor. | 30 | -.20 | -.41 | .39 | .04 |
| 11. Der Lehrer erklärt, wie er unsere Klassenarbeiten beurteilt. | 74 | .44 | .36 | .09 | .16 |
| 12. Der Lehrer gibt uns Aufgaben, ohne zu sagen, wie wir sie ausführen sollen. | 84 | .51 | -.55 | .27 | .03 |
| 13. Ich gehe gern zur Schule. | 72 | .47 | .40 | -.08 | .05 |

| | P | $r_{it}$ | Faktor I | Faktor II | Faktor III |
|---|---|---|---|---|---|
| 14. Der Lehrer bringt uns Dinge bei, die wir gerne können möchten. | 70 | .53 | .42 | -.04 | .32 |
| 15. Wenn wir etwas getan haben, was der Lehrer falsch findet, dürfen wir sagen, warum wir es getan haben. | 77 | .45 | .43 | -.01 | .34 |
| 16. Wenn man viele Fragen hat, dann beantwortet sie der Lehrer alle. | 80 | .55 | .50 | -.13 | .22 |
| 17. Auch wenn man mal frech gewesen ist, verliert der Lehrer nicht die Ruhe. | 71 | .33 | .29 | -.19 | .16 |
| 18. Der Lehrer sorgt dafür, daß alle Schüler arbeiten, so gut sie können. | 84 | .52 | .53 | .04 | .09 |
| 19. Der Lehrer weiß normalerweise, was er will. | 84 | .49 | .50 | .00 | .04 |
| 20. Bei unserem Lehrer dürfen wir offen unsere Meinung sagen. | 84 | .60 | .52 | -.24 | .35 |
| II. Strenge | | | | | |
| 1. Er achtet auf Pünktlichkeit. | 77 | .34 | .35 | .37 | -.16 |
| 2. Er hat es gern, wenn wir still auf unseren Plätzen sitzen. | 85 | .38 | .07 | .42 | -.03 |
| 3. Der Lehrer hat schlechte Laune. | 43 | .55 | -.38 | .43 | -.22 |
| 4. Er achtet auf genaue Einhaltung von Pausenzeichen. | 51 | .30 | .25 | .29 | -.04 |
| 5. Wenn der Lehrer etwas erklärt hat, sagt er, daß es nun wohl auch dem Letzten klargeworden sein müsse | 52 | .41 | .08 | .47 | .17 |

| | P | $r_{it}$ | Faktor I | II | III |
|---|---|---|---|---|---|
| 6. Der Lehrer sagt, daß er Fragen von Schülern für dumm hält. | 26 | .47 | -.25 | .40 | -.06 |
| 7. Der Lehrer stellt einige Schüler vor der Klasse bloß. | 25 | .54 | -.42 | .45 | -.14 |
| 8. Der Lehrer erteilt Befehle. | 44 | .60 | -.17 | .57 | -.12 |
| 9. Wenn wir nicht sofort tun, was der Lehrer sagt, wird er ärgerlich. | 45 | .64 | -.30 | .56 | -.20 |
| 10. Der Lehrer schimpft mit uns. | 52 | .65 | -.31 | .54 | -.23 |
| 11. Wenn man ein Schulbuch vergißt, dann schimpft der Lehrer. | 20 | .58 | -.13 | .43 | -.17 |
| 12. Wenn in der Klasse Unruhe ist, dann bestraft der Lehrer die ganze Klasse. | 37 | .50 | -.05 | .53 | -.20 |
| 13. Der Lehrer gibt Strafarbeiten auf. | 34 | .57 | -.16 | .50 | -.16 |
| III. Ausgleichendes Gewährenlassen | | | | | |
| 1. Der Lehrer läßt Vieles durchgehen. | 58 | .45 | -.15 | -.27 | .50 |
| 2. Der Lehrer unterstützt Schüler, die in der Klasse nicht besonders beliebt sind. | 65 | .59 | .05 | .05 | .67 |
| 3. Der Lehrer bespricht mit uns Schülern die Auswahl der Unterrichtsthemen. | 63 | .48 | .26 | .03 | .38 |
| 4. Der Lehrer ist auch zu den Schülern nett, die faul sind. | 62 | .51 | .30 | -.20 | .32 |
| 5. Der Lehrer lobt uns, wenn wir gute Arbeit leisten. | 75 | .54 | .33 | -.01 | .43 |

| | P | $r_{it}$ | Faktor I | Faktor II | Faktor III |
|---|---|---|---|---|---|
| 6. Wenn Schüler untereinander Streit haben, dann sagt der Lehrer, wie sie sich verhalten sollen. | 68 | .51 | .37 | .15 | .39 |
| 7. Man kann mit ihm über alles sprechen, auch wenn man etwas ausgefressen hat. | 82 | .59 | .44 | -.23 | .46 |
| 8. Wenn der Lehrer bestraft, merkt man, daß er es nicht gern tut. | 74 | .46 | .10 | -.10 | .42 |
| 9. Der Lehrer sieht ein, daß die Schüler recht haben und gibt nach. | 67 | .55 | .31 | -.17 | .41 |
| 10. Wenn man zu spät zur Schule kommt, dann hat der Lehrer dafür Verständnis. | 77 | .45 | .09 | -.21 | .40 |
| 11. Bei unserem Lehrer können wir uns viel herausnehmen. | 51 | .42 | -.17 | -.16 | .48 |

### 4.3.2. Aufgabenanalyse und Reliabilität der Subskalen

Die drei (faktorenanalytisch gewonnenen) Subskalen wurden anschließend itemanalysiert und auf ihre Reliabilität hin überprüft. Tabelle 2 gibt neben den Faktorenladungen die Schwierigkeitsindizes und Trennschärfekoeffizienten $r_{it}$ der Items wieder. Da mit einer vierstufigen Antwortskala gearbeitet wurde, mußte der Schwierigkeitsindex über $P = (\bar{X} - 1)$ mal 100 : 3 zurückgerechnet werden. Als Beantwortungsmodus der Items wurden die in der Untersuchung von BARTRAM (1972) verwendeten Antwortalternativen beibehalten, wie z.B. bei Item 1:

Der Lehrer freut sich, wenn er die Klasse betritt.
oft manchmal selten nie

Außer Item 9 und 10 der Skala Unterstützung erweisen sich

alle weiteren Items des Fragebogens als trennscharf. Die Schwierigkeitsindizes sind für Strenge und Ausgleichendes Gewährenlassen hinreichend ausbalanciert. Die Reliabilitätsschätzungen der drei Skalen erfolgte mittels der Minimalschätzung nach Kuder-Richardson (vergl. LIENERT, 1967). Sie betragen für

| | |
|---|---|
| UNTERSTÜTZUNG | 0,70 |
| STRENGE | 0,74 |
| AUSGLEICHENDES GEWÄHRENLASSEN | 0,73 |

Die Reliabilitätsschätzung für Unterstützung dürfte höher ausfallen, wenn die beiden unbrauchbaren Items 9 und 10 in die Berechnung nicht eingehen.

Um die Stabilität der Schülerbeurteilung für einzelne Lehrer über die Zeit zu ermitteln, wurde nach drei Monaten mit vier Gesamtschulklassen eine Testwiederholung vorgenommen. Korreliert wurden die Itemmittelwerte jeder Skala für jede der vier Klassen. Tabelle 3 zeigt die Retest-Koeffizienten.

Tabelle 3

Retest-Koeffizienten der drei Skalen

| | UNTERSTÜTZUNG | STRENGE | AUSGLEICHENDES GEWÄHRENLASSEN |
|---|---|---|---|
| Klasse 1 | .68 | .92 | .96 |
| Klasse 2 | .78 | .97 | .93 |
| Klasse 3 | .81 | .89 | .85 |
| Klasse 4 | .75 | .97 | .64 |
| $\bar{r}$ | .76 | .95 | .89 |

Tab. 4

Punktwerte für die Real-, Schätz- und Wunschurteile der Lehrer bezüglich der drei Skalen

| Lehrer Nr. | Schulart | Real-urteile (UNTERSTÜTZUNG) | Schätz-urteile (UNTERSTÜTZUNG) | Wunsch-urteile (UNTERSTÜTZUNG) | Real-urteile (STRENGE) | Schätz-urteile (STRENGE) | Wunsch-urteile (STRENGE) | Real-urteile (AUSGLEICHENDES GEWÄHRENLASSEN) | Schätz-urteile (AUSGLEICHENDES GEWÄHRENLASSEN) | Wunsch-urteile (AUSGLEICHENDES GEWÄHRENLASSEN) |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 ♂ | HS | 4,3 | - | - | 17,9 | - | - | 12,5 | - | - |
| 2 ♀ | HS | 15,1 | 18,0 | 26,0 | 20,8 | 21,0 | 12,0 | 6,7 | 12,0 | 20,0 |
| 3 ♂ | HS | 13,0 | 1,0 | 26,0 | 13,6 | 20,0 | 19,0 | 10,4 | 1,0 | 10,0 |
| 4 ♂ | HS | 16,1 | - | - | 16,5 | - | - | 10,1 | - | - |
| 5 ♂ | HS | 19,1 | 9,0 | 27,0 | 14,1 | 18,0 | 5,0 | 16,4 | 12,0 | 19,0 |
| 6 ♀ | HS | 16,7 | 9,0 | 29,0 | 12,5 | 16,0 | 3,0 | 17,7 | 17,0 | 24,0 |
| 7 ♂ | HS | 22,0 | 19,0 | 24,0 | 9,9 | 12,0 | 9,0 | 13,5 | 5,0 | 16,0 |
| 8 ♂ | HS | 19,2 | 14,0 | 22,0 | 8,3 | 17,0 | 15,0 | 13,1 | 21,0 | 21,0 |
| 9 ♀ | HS | 18,0 | 17,0 | 23,0 | 11,6 | 15,0 | 12,0 | 14,1 | 17,0 | 21,0 |
| 10 ♂ | HS | 13,6 | 16,0 | 24,0 | 11,5 | 10,0 | 3,0 | 16,2 | 19,0 | 23,0 |
| 11 ♀ | HS | 20,1 | 25,0 | 21,0 | 16,0 | 11,0 | 14,0 | 14,9 | 18,0 | 17,0 |
| 12 ♂ | HS | 22,4 | - | - | 9,8 | - | - | 18,0 | - | - |
| | $\bar{x}$ | 17,4 | 14,2 | 24,7 | 13,1 | 15,6 | 10,2 | 13,7 | 13,6 | 18,4 |
| | s | 2,9 | 6,1 | 2,4 | 3,5 | 3,7 | 5,3 | 3,2 | 6,4 | 3,8 |
| 13 ♀ | GS | 14,1 | 17,0 | 26,0 | 15,2 | 20,0 | 14,0 | 8,0 | 13,0 | 17,0 |
| 14 ♂ | GS | 15,3 | 20,0 | 26,0 | 16,2 | 12,0 | 9,0 | 10,7 | 17,0 | 18,0 |
| 15 ♂ | GS | 21,0 | 23,0 | 28,0 | 9,1 | 7,0 | 4,0 | 14,9 | 17,0 | 14,0 |
| 16 ♂ | GS | 19,0 | 24,0 | 29,0 | 11,8 | 13,0 | 9,0 | 15,0 | 18,0 | 19,0 |
| 17 ♀ | GS | 21,0 | 15,0 | 31,0 | 8,1 | 12,0 | 7,0 | 15,3 | 13,0 | 17,0 |
| 18 ♂ | GS | 17,6 | 15,0 | 27,0 | 12,5 | 12,0 | 3,0 | 12,3 | 10,0 | 19,0 |
| 19 ♀ | GS | 14,1 | 9,0 | 21,0 | 16,2 | 20,0 | 12,0 | 8,4 | 4,0 | 15,0 |
| 20 ♀ | GS | 17,9 | 11,0 | 22,0 | 16,3 | 19,0 | 16,0 | 10,3 | 8,0 | 16,0 |
| 21 ♂ | GS | 19,3 | 23,0 | 24,0 | 10,6 | 5,0 | 9,0 | 15,6 | 19,0 | 21,0 |
| 22 ♂ | GS | 18,2 | 14,0 | 29,0 | 12,7 | 19,0 | 1,0 | 12,3 | 13,0 | 23,0 |
| 23 ♂ | GS | 15,1 | 13,0 | 27,0 | 13,8 | 8,0 | 3,0 | 11,1 | 20,0 | 22,0 |
| 24 ♀ | GS | 22,1 | 16,0 | 26,0 | 11,8 | 12,0 | 6,0 | 14,8 | 16,0 | 18,0 |
| 25 ♂ | GS | 22,2 | 19,0 | 27,0 | 7,7 | 7,0 | 2,0 | 16,8 | 11,0 | 19,0 |
| 26 ♂ | GS | 22,3 | 15,0 | 19,0 | 3,8 | 6,0 | 7,0 | 17,1 | 12,0 | 18,0 |
| | $\bar{x}$ | 18,5 | 16,9 | 25,9 | 11,8 | 12,3 | 7,3 | 13,0 | 13,6 | 18,0 |
| | s | 2,9 | 4,5 | 3,2 | 3,6 | 5,5 | 4,4 | 2,9 | 4,4 | 2,4 |
| | $\bar{x}$ (gesamt) | 18,1 | 15,7 | 25,4 | 12,4 | 13,6 | 8,4 | 13,3 | 13,4 | 18,3 |

( HS = Hauptschule; GS = Gesamtschule )

Die Koeffizienten erwiesen sich als nicht heterogen, deswegen konnte eine Mittelung über z' vorgenommen werden.

Die Punktwerte der Real-, Schätz- und Wunschurteile der an der Untersuchung beteiligten Lehrer sind in Tabelle 4 aufgeführt, ebenfalls Mittelwerte und Streuungen.

Lehrer 1, 4 und 12 gaben keine Schätz- und Wunschurteile ab. Die Punktwerte der Realurteile dieser Lehrer bleiben bei der weiteren Verrechnung der Daten unberücksichtigt.

### 4.3.3. Varianzanalytische und korrelationsstatistische Auswertung des Zahlenmaterials der Tabelle 4 im Hinblick auf die Fragestellungen

Die eigentlichen Fragestellungen dieser Voruntersuchung lassen sich mit einer mehrfaktoriellen Varianzanalyse beantworten.

Die Varianzanalyse ist ein statistisches Prüfverfahren, das Unterschiede zwischen Gruppen von normal verteilten, homogen varianten Daten (Meßwerten), mindestens auf Intervallskalenniveau hinsichtlich ihrer Mittelwerte untersucht. Dabei kann die Gruppierung des Datenmaterials nach nur einem, nach zwei oder nach mehreren unabhängigen Variablen, in unserem Fall nach Schultyp, Lehrergeschlecht und Beurteilungsmodus, mit oder ohne Berücksichtigung der möglichen Wechselwirkungen zwischen den unabhängigen Variablen (Faktoren) erfolgen.

Zur Überprüfung von signifikanten Haupteffekten zwischen den Stufen des Beurteilungsmodus (erhaltenen, geschätzten und erwünschten Punktwerten bezüglich der drei Skalen) und signifikanten Wechselwirkungseffekten zwischen dem Beurteilungsmodus und dem Schultyp einerseits un dem Beurteilungsmodus und dem Lehrergeschlecht andererseits wird ein two-factor-mixed design verwendet (vergl. BRUNING und KINTZ, 1968, und WINER, 1970).

Unterstützung, Strenge und Ausgleichendes Gewährenlassen stellen die abhängigen Variablen dar. Als unabhängige Variablen haben wir:

1. Beurteilungsmodus (auf den Stufen real, geschätzt, erwünscht); diese Variable ist ein random factor, da es sich um abhängige Messungen (Meßwiederholungen pro Lehrer) handelt.

2. Schultyp (Hauptschule, Gesamtschule)

3. Lehrergeschlecht

4. Lehrer

Für jede der drei abhängigen Variablen (Unterstützung, Strenge und Ausgleichendes Gewährenlassen) rechnen wir zwei Varianzanalysen, also insgesamt sechs.
In die erste Varianzanalyse gehen folgende Variablen ein:

- Abhängige Variable: Unterstützung
- Unabhängige Variable: Schultyp (Versuchsbedingungen)
- Unabhängige Variable: Beurteilungsmodus (Versuchswiederholungen)
- Unabhängige Variable: Lehrer (Vpn)

In der zweiten Varianzanalyse tritt anstelle des Faktors Schultyp das Lehrergeschlecht.

Als abhängige Variablen gehen in die dritte und vierte Varianzanalyse Strenge, in die fünfte und sechste Ausgleichendes Gewährenlassen ein. Analog zur ersten und zweiten Varianzanalyse werden in der dritten und vierten sowie fünften und sechsten die unabhängigen Variablen beibehalten. Tabelle 5 a zeigt die Ergebnisse der sechs Varianzanalysen. Hierbei bezeichnet

Zw. Vpn Varianz zwischen den Versuchspersonen (Lehrern)
Bedingungen: Varianz zwischen den Versuchsbedingungen (Schultyp bzw. Lehrergeschlecht)

Fehler (zw.) : Interindividuelle Fehlervarianz
Innerhalb Vpn. : Varianz innerhalb der Vpn (Lehrer)
Beurteilungsmodus : Varianz aufgrund der Versuchswiederholungen (d.h. zwischen den Stufen des Beurteilungsmodus)
Modus x Bedingungen: Wechselwirkungsvarianz
Fehler (inn.) : Fehlervarianz innerhalb der Vpn (Intraindividuelle Fehlervarianz)

Tabelle 5 a

Ergebnisse der Varianzanalysen

1. Varianzanalyse (abhängige Variable = Unterstützung)

| QUELLE | QS | FG | MQS | F-TEST | |
|---|---|---|---|---|---|
| Total | 2263.00 | 68 | -- | -- | |
| Zw. Vpn. | 471.00 | 22 | -- | -- | |
| Bedingungen (Schultyp) | 42.00 | 1 | 42.00 | 2.15 | |
| Fehler (zw.) | 429.00 | 21 | 19.50 | -- | |
| Innerhalb Vpn | 1792.00 | 46 | -- | -- | |
| Beurteilungsmodus | 1166.00 | 2 | 583.00 | 47.00 | **F 0,99 (5,1 |
| ModusxBedingungen | 105.00 | 2 | 52.50 | 4.20 | *F 0,95 (3,2 |
| Fehler (inn.) | 521.00 | 42 | 12.40 | -- | |

2. Varianzanalyse (abhängige Variable = Unterstützung)

| QUELLE | QS | FG | MQS | F-TEST | |
|---|---|---|---|---|---|
| Total | 2264.00 | 68 | -- | -- | |
| Zw. Vpn. | 471.00 | 22 | -- | -- | |
| Bedingungen (Schultyp) | 8.00 | 1 | 8.00 | 0.36 | |
| Fehler (zw.) | 463.00 | 21 | 22.00 | -- | |
| Innerhalb Vpn | 1793.00 | 46 | -- | -- | |
| Beurteilungsmodus | 1165.00 | 2 | 582.50 | 39.10 | **F 0,99(5,15 |
| ModusxBedingungen | 1.00 | 2 | 0.50 | 0.02 | |
| Fehler (inn.) | 627.00 | 42 | 14.90 | -- | |

3. Varianzanalyse (abhängige Variable = Strenge)

| QUELLE | QS | FG | MQS | F-TEST |
|---|---|---|---|---|
| Total | 1752.00 | 68 | -- | -- |
| Zw. Vpn. | 921.00 | 22 | -- | -- |
| Bedingungen (Schultyp) | 103.00 | 1 | 103.00 | 2.35 |
| Fehler (zw.) | 818.00 | 21 | 43.80 | -- |
| Innerhalb Vpn | 831.00 | 46 | -- | -- |
| Beurteilungsmodus | 336.00 | 2 | 168.00 | 8.47 **F 0,99 (5,15) |
| ModusxBedingungen | 6.00 | 2 | 3.00 | 0.15 |
| Fehler (inn.) | 489.00 | 42 | 19.80 | -- |

4. Varianzanalyse (abhängige Variable = Strenge)

| QUELLE | QS | FG | MQS | F-TEST |
|---|---|---|---|---|
| Total | 1752.00 | 68 | -- | -- |
| Zw. Vpn. | 949.00 | 22 | -- | -- |
| Bedingungen (Geschlecht) | 229.00 | 1 | 229.00 | 6.70 *F 0,95 (4,32) |
| Fehler (zw.) | 720.00 | 21 | 34.30 | -- |
| Innerhalb Vpn | 803.00 | 46 | -- | -- |
| Beurteilungsmodus | 331.00 | 2 | 165.50 | 14.90 **F 0,99(5,15) |
| ModusxBedingungen | 4.00 | 2 | 2.00 | 0.20 |
| Fehler (inn.) | 468.00 | 42 | 11.10 | -- |

5. Varianzanalyse (abhängige Variable = Ausgleichendes Gewährenlassen)

| QUELLE | QS | FG | MQS | F-TEST |
|---|---|---|---|---|
| Total | 477.00 | 68 | -- | -- |
| Zw. Vpn. | 673.00 | 22 | -- | -- |
| Bedingungen (Schultyp) | 0.00 | 1 | 0.00 | 0.00 |
| Fehler (zw.) | 673.00 | 21 | 32.00 | -- |
| Innerhalb Vpn | 804.00 | 46 | -- | -- |
| Beurteilungsmodus | 378.00 | 2 | 189.00 | 19.30 **F 0,99 (5,15) |
| ModusxBedingungen | 14.00 | 2 | 7.00 | 0.71 |
| Fehler (inn.) | 412.00 | 42 | 9.80 | -- |

6. Varianzanalyse (abhängige Variable = Ausgleichendes Gewährenlassen)

| QUELLE | QS | FG | MQS | F-TEST | |
|---|---|---|---|---|---|
| Total | 477.00 | 68 | -- | -- | |
| Zw. Vpn. | 994.00 | 22 | -- | -- | |
| Bedingungen (Geschlecht) | 31.00 | 1 | 31.00 | 0.70 | |
| Fehler (zw.) | 963.00 | 21 | 45.00 | -- | |
| Innerhalb Vpn. | 483.00 | 46 | -- | -- | |
| Beurteilungsmodus | 378.00 | 2 | 189.00 | 75.60 | **F 0,99 (5,15) |
| ModusxBedingungen | 1.00 | 2 | 0.50 | 0.20 | |
| Fehler (inn.) | 104.00 | 42 | 2.50 | -- | |

Die Ergebnisse der Varianzanalysen weisen hochsignifikante Haupteffekte zwischen den Beurteilungsmodi bezüglich aller drei Skalen aus.

Mit dem t-Test für Paardifferenzen (vergl. HOFSTÄTTER, 1966, S. 101) ermitteln wir unter Zugrundelegung einer einseitigen Fragestellung zwischen Real- und Schätzurteilen für Unterstützung ein auf dem 5% - Niveau signifikanten $t_{corr}$- Wert von 2,21; für Strenge erhalten wir zwischen Real- und Schätzwerten ein $t_{corr}$ von 1,56 ($p < 0,10$). Vergl. hierzu die Mittelwerte in Tabelle 4.

Hinsichtlich Unterstützung (1. VA) finden wir signifikante Wechselwirkungseffekte ($p < 0,05$) zwischen den unabhängigen Variablen Schultyp und Beurteilungsmodus; d.h., die Gesamtschullehrer schätzen, wünschen und erhalten höhere Unterstützungsscores als die Hauptschullehrer.

Hinsichtlich Strenge wird die Wechselwirkungsvarianz zwischen Beurteilungsmodus und Lehrergeschlecht (4. VA)

signifikant ($p < 0,05$), d.h., daß die Lehrerinnen in unserer Stichprobe höhere Strengescores schätzen, wünschen und erhalten.

Zusätzlich zu den Ergebnissen der Varianzanalysen errechneten wir zwischen allen Beurteilungsmodi folgende Korrelationskoeffizienten:

Tabelle 5 b

Korrelationsmatrix der Beurteilungsmodi

| | U Real | U Wu | U Schä | S Real | S Wu | S Schä | A Real | A Wu | A Schä |
|---|---|---|---|---|---|---|---|---|---|
| U Real | - | -0,02 | 0,44 | -0,65 | -0,28 | -0,53 | 0,63 | -0,06 | 0,11 |
| U Wu | | - | 0,01 | -0,02 | -0,49 | 0,04 | 0,12 | 0,25 | 0,16 |
| U Schä | | | - | -0,12 | -0,17 | -0,56 | 0,19 | 0,04 | 0,50 |
| S Real | | | | - | 0,35 | 0,62 | -0,73 | -0,13 | -0,19 |
| S Wu | | | | | - | 0,49 | -0,54 | -0,58 | -0,35 |
| S Schä | | | | | | - | -0,62 | -0,10 | -0,42 |
| A Real | | | | | | | - | 0,27 | 0,37 |
| A Wu | | | | | | | | - | 0,62 |
| A Schä | | | | | | | | | - |

U = Unterstützung
S = Strenge
A = Ausgleichendes Gewährenlassen
Real = Realurteil
Wu = Wunschurteil
Schä = Schätzurteil

Zwecks besserer Übersichtlichkeit reduzieren wir mittels Hauptkomponentenanalyse diese Korrelationsmatrix auf folgende Faktorenmatrix:

Tabelle 5 c
Faktorenmatrix der Urteilsmodi

| Urteilsmodus | Realfaktor I | Wunschfaktor II | Schätzfaktor III |
|---|---|---|---|
| 1) U Real | 0,77 | -0,03 | 0,20 |
| 2) U Wu | 0,06 | 0,57 | -0,11 |
| 3) U Schä | 0,23 | -0,11 | 0,80 |
| 4) S Real | -0,83 | -0,16 | -0,05 |
| 5) S Wu | -0,48 | -0,65 | -0,16 |
| 6) S Schä | -0,69 | -0,10 | -0,51 |
| 7) A Real | 0,81 | 0,33 | 0,16 |
| 8) A Wu | 0,05 | 0,63 | 0,31 |
| 9) A Schä | 0,14 | 0,36 | 0,74 |

Die drei erhaltenen Faktoren lassen sich eindeutig als "Real-, Wunsch- und Schätzfaktor" interpretieren.

Außer den Realurteilen laden die gewünschte und geschätzte Strenge noch relativ hoch auf dem "Realfaktor". Bezüglich Strenge kovariiert - im Vergleich zu Unterstützung und Ausgleichendem Gewährenlassen - das Schätzurteil mit dem Realurteil und dem Wunschurteil relativ hoch; d.h. Lehrer, die hohe Strengeurteile von ihren Schülern vermuten, werden von ihren Schülern auch als streng beurteilt und wünschen sich zugleich hohe Strengeurteile.

Darüberhinaus fällt die durchschnittliche Korrelation zwischen Real-, Schätz- und Wunschurteil für Strenge mit $r \approx 0{,}50$ höher aus als für Unterstützung und Ausgleichendes Gewährenlassen ($r \approx 0{,}30$). Die Korrelationskoeffizienten wurden über z' gemittelt.

## 4.4. Zusammenfassung der Ergebnisse der Voruntersuchung

- Bezüglich Unterstützung (U), Strenge (S) und Ausgleichendes Gewährenlassen erhalten wir signifikante Unterschiede zwischen den Real-, Schätz- und Wunschurteilen:

  1) U Schätzurteile < U Realurteile < U Wunschurteile
  2) S Wunschurteile < S Realurteile < S Schätzurteile
  3) A Realurteile und A Schätzurteile < A Wunschurteile

- Gesamtschullehrer schätzen, wünschen und erhalten höhere Unterstützungsscores als Hauptschullehrer.

- Lehrerinnen schätzen, wünschen und erhalten höhere Strengescores als Lehrer.

- Für Strenge kovariieren Real-, Schätz- und Wunschurteile im Durchschnitt höher als für Unterstützung und Ausgleichendes Gewährenlassen. Insbesondere kovariieren (hinsichtlich Strenge) das Schätzurteil mit dem Realurteil einerseits und das Schätzurteil mit dem Wunschurteil andererseits relativ hoch.

## 4.5. Diskussion der Ergebnisse der Voruntersuchung

Ein zweidimensionales Modell elterlicher Bekräftigungsweisen (Strenge und Unterstützung), wie es HERRLICHN et al. (1971) zur Erfassung des elterlichen Erziehungsstils verwenden, erwies sich - modifiziert auf die Lehrer - Schüler - Interaktion - auch zur Charakterisierung des Lehrerverhaltens aus der Sicht der Schüler als brauchbar. Neben Strenge und Unterstützung erbrachte die Faktorenanalyse der Items noch eine Dimension, die wir als Ausgleichendes Gewährenlassen interpretierten, weil auf ihr Items laden, in denen, im Vergleich zu gezielter und geordneter Unterstützung, eher Nachgiebigkeit und Ver-

ständnis den Schülern gegenüber zum Ausdruck kommen. Die vom Lehrer erlebte Unterstützung (Faktor I) der Schüler bedeutet u.a., daß die Kinder ihrem Urteil zufolge vom Lehrer gezielte Hilfe und Förderung, verbunden mit freundlicher Zuwendung, erhalten. Erlebte Strenge (Faktor II) bedeutet u.a., daß die Schüler über Tadel, Blamage bei Leistungsversagen, Strafen, Schimpfen und Befehle des Lehrers berichten.

Die Trennschärfekoeffizienten in der Größenordnung von 0,37 - 0,60 für Unterstützung (wenn man von Item 9 und 10 absieht), von 0,34 - 0,65 für Strenge und 0,42 - 0,59 für Ausgleichendes Gewährenlassen, sowie die für die drei Dimensionen erhaltenen Reliabilitätsschätzungen von 0,70, 0,74 und 0,73 sprechen für die Brauchbarkeit des Itempools. Die hohen Retest-Koeffizienten (0,76 für Unterstüzung, 0,95 für Strenge und 0,89 für Ausgleichendes Gewährenlassen) weisen darauf hin, daß relativ stabile Verhaltensmerkmale über die Zeit erfaßt wurden, wenngleich die Wiederholungszuverlässigkeit für Unterstützung von 0,76 etwas abfällt. Hier zeigt sich, worauf im Kapitel 2.2.3. bereits hingewiesen wurde, der Vorteil der Methode der Verhaltenseinschätzung von Lehrern durch Schüler, nämlich daß die Reliabilität der Schätzurteile von Schülern über Lehrer im Sinne der SPEARMAN-BROWN-Formel eine Funktion der Anzahl der Beurteiler ist. Da sich die Itemmittelwerte bei ca. 30 und mehr Schülern pro Klasse stabilisieren, werden die Schätzurteile so reliabel wie Schulleistungs- und Intelligenztests.

Die Voruntersuchung erbrachte außerdem, daß die Lehrer von ihren Schülern als unterstützender beurteilt werden als sie vermuten. Darüberhinaus deutet sich als Trend an, daß Schüler ihre Lehrer weniger streng einstufen als die Lehrer vermuten. Betrachtet man ein hohes Ausmaß an Unterstützung und ein geringes Ausmaß an Strenge als "angebrachtes" Interaktionsverhalten des Lehrers, so läßt sich

zusammenfassend feststellen, daß Lehrer von ihren Schülern bezüglich ihres Interaktionsstils als "günstiger" wahrgenommen werden als sie annehmen. Demnach neigen die in der Voruntersuchung beurteilten Lehrer dazu, "ihr Licht unter den Scheffel zu stellen". Für dieses überraschende Ergebnis, das durch die Hauptuntersuchung noch einmal überprüft werden soll, bieten sich zwei Interpretationsalternativen an:

- In der Erwartung, daß die Schüler den Lehrern an Unterstützung einen Mangel und an Strenge ein Zuviel bescheinigen könnten, entschärfen die Lehrer im voraus eine vermeintliche Mißerfolgsrückmeldung, indem sie sich weniger Unterstützung und mehr Strenge im Urteil der Schüler zuschreiben. Aus Furcht, von den Schülern ungünstig beurteilt zu werden, möchte man die Diskrepanz zwischen Realurteilen und geschätzten Realurteilen nicht zu groß werden lassen.

- Als zweite Interpretationsmöglichkeit könnte sich ein in der Hauptuntersuchung zu kontrollierendes response set andeuten. Es sei die Vermutung geäußert, daß die Mittelwertsunterschiede zwischen Schätz- und Realurteilen durch die Wirkung "sozialer Erwünschtheit" zustande gekommen sein könnten, d.h., daß die Lehrer von den Schülern hinsichtlich Unterstützung höhere und hinsichtlich Strenge niedrigere Werte aufgrund sozial erwünschter Antworttendenzen der Schüler erhalten. Deshalb soll das Itemmaterial in der Hauptuntersuchung u.a. daraufhin überprüft werden, ob sich in bestimmten Items durch die Schülerantworten die Tendenz zur sozialen Erwünschtheit manifestiert.

Weiterhin stellte sich in unserer Untersuchung heraus, daß die von Lehrern erwünschten Schülerurteile erheblich von den tatsächlich erhaltenen Schülerurteilen abweichen, derart, daß die Lehrer unterstützender, ausgleichender und weniger streng sein möchten. Dieses Ergebnis und die Tat-

sache, daß für Strenge die Schätz- und Wunschurteile und insbesondere die Real- und Schätzurteile höher kovariieren als für Unterstützung und Ausgleichendes Gewährenlassen, erscheinen nicht überraschend, wenn man in Rechnung stellt, daß Strenge und Dirigismus die von Lehrern weitaus am häufigsten praktizierten Verhaltensweisen darstellen. Wenn man die Faktoren Unterstützung und Ausgleichendes Gewährenlassen im Lehrerverhalten als wünschenswert ansieht und hier die Real- und Schätzurteile weiter auseinanderfallen als bei Strenge, wäre es neben einer ständigen Rückmeldung in diesen Dimensionen wichtig, die Lehrer in den Techniken der Verhaltensmodifikation, speziell in der Handhabung der positiven sozialen Verstärker, zu trainieren. Vergl. hierzu u.a. die Arbeiten von MADSEN et al. (1968), CURRY (1970), WETZEL (1970) und WHITMAN (1971).

Die signifikante Wechselwirkung zwischen Schultyp und Beurteilungsmodus bezüglich Unterstützung könnte durch einen Selektionseffekt zustande gekommen sein. Die Gesamtschüler setzen sich in unserer Stichprobe vorwiegend aus jüngeren, die Hauptschüler aus älteren Kindern zusammen. Wir vermuten, daß ältere Schüler ihre Lehrer kritischer beurteilen als jüngere Schüler. Diese Annahme legt eine Untersuchung von GRUBITSCH und VOGT (1970) nahe, in der nachgewiesen werden konnte, daß jüngere Schüler eine größere Gemeinsamkeit mit ihrem Wunschlehrerbild zeigen als dies bei älteren Schülern der Fall ist. Um diesen mutmaßlichen Alterseffekt zu reduzieren, ist für die Hauptuntersuchung an eine altershomogenere Stichprobe gedacht.

Eine zweite signifikante Wechselwirkung zeigte sich zwischen Lehrergeschlecht und Beurteilungsmodus. Demnach schätzen, wünschen und erhalten Lehrerinnen höhere Strengescores als Lehrer. Da wir annehmen, daß es sich hier um ein stichprobenspezifisches Ergebnis handelt, möchten wir auf eine Erörterung dieses Ergebnisses verzichten.

## III. ERSTER TEIL DER HAUPTUNTERSUCHUNG

### 5. Vorbemerkungen zum Ablauf der Hauptuntersuchung

Während die Voruntersuchung im wesentlichen Aufschlüsse über die generelle Brauchbarkeit der Items zur Charakterisierung von Lehrerverhalten aus der Sicht der Schüler erbringen sollte, ist es das wichtigste Anliegen des ersten Teils der Hauptuntersuchung, durch Itemselektion eine nach Möglichkeit intelligenzunabhängige Fragebogenendform zu erstellen; d.h. die endgültige Erziehungsskala zur Erfassung angemessenen Lehrer-Interaktionsverhaltens soll solche Items enthalten, die auch Verhalten von Sonderschullehrern aus der Sicht von Schülern in Lernbehindertenklassen noch hinreichend reliabel messen. Aus diesem Grunde bezieht die Hauptuntersuchung neben Real- und Hauptschülern auch lernbehinderte Schüler ein.

Der zweite Teil der Hauptuntersuchung wird Validierungsaspekte umfassen. Gesucht werden Lehrergruppen, die sich bezüglich erwünschter, geschätzter und tatsächlich erhaltener Schülerurteile unterscheiden. Insbesondere wird, unter Zugrundelegung der HOFERSCHEN Urteilsfaktoren zur aktuellen Schülerbeurteilung, nach Lehrerbeurteilertypen Ausschau gehalten. Diese Lehrerbeurteilertypen werden hinsichtlich ihrer Diskrepanzen zwischen Real- und Schätzurteilen miteinander verglichen.

In der Voruntersuchung stellte die Faktorenanalyse ein provisorisches Hilfsmittel dar, um Aufschluß über die Dimensionalität des Itempools zu erhalten. In der Hauptuntersuchung wird zunächst nachzuprüfen sein, ob sich die Dreierstruktur (Unterstützung, Strenge und Ausgleichendes Gewährenlassen) als stabil erweist. Im Anschluß daran - vor der eigentlichen Itemselektion - wird noch einmal anhand einer detaillierten Auswertung das aus der Vorunter-

suchung herausragende Ergebnis kontrolliert werden, daß Lehrer das Ausmaß ihrer von den Schülern perzipierten Strenge besser einzuschätzen in der Lage sind als das Ausmaß an Unterstützung.

## 6. Fragestellungen der Hauptuntersuchung

1) Läßt sich die in der Voruntersuchung erhaltene Dimensionalität des Itempools - die als Unterstützung, Strenge und Ausgleichendes Gewährenlassen interpretierte Dreierstruktur - auch in der Hauptuntersuchung reproduzieren, in die eine aus Haupt-, Real- und Sonderschülern bestehende, relativ altershomogene Schülerpopulation eingeht?

2) Erweist sich der Itempool - nach aufgabenanalytischen Kriterien - auch für Sonderschulklassen als brauchbar?

3) Können die in die Hauptuntersuchung einbezogenen Lehrer das Ausmaß ihrer von den Schülern perzipierten Strenge besser einschätzen als das Ausmaß an Unterstützung?

4) Welche Items eignen sich, einem Expertenrating zufolge, am besten zur Charakterisierung a n g e b r a c h t e n Interaktionsverhaltens von Lehrern?

5) Welchte Items korrelieren gering mit sozial erwünschtem Antwortverhalten?

6) Welche Items weisen innerhalb von Klassen geringe und zwischen Klassen große Beurteilungsunterschiede auf?

7) Lassen sich bezüglich angemessenen Interaktionsverhaltens im Unterricht bei den an der Untersuchung beteiligten Lehrern schultypen- und geschlechtsspezifische Unterschiede nachweisen?

8) Unterscheiden sich besonders günstig beurteilte Lehrer im Vergleich zu eher ungünstig beurteilten Lehrern in ihren Schätzurteilen (vermuteten Realurteilen)?

9) Weisen von ihren Schülern günstig eingestufte Lehrer im Vergleich zu eher ungünstig eingestuften Lehrern unterschiedliche Diskrepanzen zwischen
a) ihren Real- und Wunschurteilen und
b) ihren Schätz- und Wunschurteilen auf?

10) Lassen sich aufgrund der HOFERSCHEN Beurteilungsdimensionen zur aktuellen Schülerbeurteilung unterschiedliche Lehrerbeurteilertypen finden?

11) Bei welchen Lehrerbeurteilertypen weichen die geschätzten (vermuteten) Schülerurteile von den tatsächlich erhaltenen Schülerurteilen in auffälligem Ausmaß ab?

12) Neigen günstig und ungünstig eingestufte Lehrer zu unterschiedlichen Schülerbeurteilungen?

## 7. Durchführung der Hauptuntersuchung

### 7.1. Untersuchungsmaterial

#### 7.1.1. Die Items zur Lehrerbeurteilung für die Hauptuntersuchung

Bis auf Item 9 und 10 der Unterstützungsdimension, die niedrige negative Trennschärfen aufweisen (vergl. Tabelle 2), werden alle Items aus der Voruntersuchung beibehalten. Zusätzlich fügen wir noch folgende Unterstützungsitems hinzu:
- Der Lehrer kann mich ganz gut leiden.
- Der Lehrer ist zu mir gerecht.
- Zu mir ist der Lehrer höflich und nett.

Als Strengeitems werden noch aufgenommen:
- Der Lehrer wird leicht ärgerlich.
- An mir hat der Lehrer etwas auszusetzen.
- Der Lehrer ist streng.

Ein Exemplar des Schülerfragebogens zur Lehrerbeurteilung ist, einschließlich Instruktion, dem Anhang beigefügt.

### 7.1.2. Fragebogen zur Erfassung von Simulationstendenzen

Kontrollinstrumente zur Messung sozialer Erwünschtheit oder Lügenlisten, wie Psychologen im Fachjargon sagen, werden zur Entlarvung von (Dis)Simulationen entwickelt. "Sie bestehen aus Fragen, welche von normalen Personen im allgemeinen in bejahendem Sinne beantwortet werden. Eine übermäßige Ablehnung solcher Lügenfragen deutet im allgemeinen auf Dissimulation hin und übermäßige Zustimmung im allgemeinen auf Simulation" (EYSENCK, 1964, S.3

Nach GUILFORD (1965, S. 172) basieren Lügenskalen auf "Antworten zu Feststellungen, die moralisch wertvolle Dinge behaupten, die kaum auf jemanden zutreffen." Simulationstendenzen hängen von der jeweiligen Bezugssituation und auch -person ab, in der sich Probanden befinden. Kinder und Jugendliche werden sich in der Schule und dem Lehrer gegenüber anders verhalten und damit auch anders bei Befragungen antworten als beispielsweise zu Hause oder in Gesprächen mit Spielgefährten. Um zu kontrollieren, inwieweit unsere Items zur Lehrerbeurteilung durch sozial erwünschte Antworttendenzen der Schüler determiniert sind, bot sich ein von ASCHERSLEBEN (1971) entwickelter Lügenscor zur Messung von Simulationstendenzen an. Dieser Lügenscore ist speziell für den schulischen Bereich geeignet. Die Reliabilität und inhaltliche Homogenität dieses Kontrollinstrumentes sind hinreichend hoch ($r_{stab}$ = 0,84). Volks-

und Oberschüler unterscheiden sich hochsignifikant, wobei Volksschüler stärker zum Simulieren neigen als Oberschüler. Vermutlich erhalten Sonderschüler (Lernbehinderte) den höchsten mittleren Lügenscore. Der Lügenfragebogen ist im Anhang zu finden.

### 7.1.3. Die Urteilseigenschaften für die Lehrer zur aktuellen Schülerbeurteilung

Im täglichen Umgang geben wir ständig Urteile über andere Menschen ab. Wir stufen unsere Mitmenschen als aufrichtig, frech, verträumt, unstet, willig, kameradschaftlich, ungeduldig, aufdringlich, fleißig, redselig usw. ein. Ein Arzt spricht beispielsweise über seine Patienten; die dabei abgegebenen Stellungnahmen sind unreflektiert, d.h. nicht rational begründet. Hat man mit einem Menschen zu tun, dessen Gewohnheiten einem überhaupt nicht gefallen, so neigt man leicht dazu, ihn als unsympathisch zu charakterisieren. Alle diese Urteile sind unreflektiert, sie liegen unserem Verhalten dem Mitmenschen gegenüber zugrunde. Die Frage, ob diese Urteile "richtig" sind, ist ein vom wissenschaftlichen Standpunkt aus unangemessener Zugang zum angesprochenen Problem. Stattdessen soll die Menschenbeurteilung als Teilgegenstand s o z i a l e n V e r h a l t e n s betrachtet werden, gewissermaßen als kognitive und konativeKomponente des sozialen Verhaltens. Wenn man die Frage nach der Richtigkeit stellt, wäre es kein soziales Verhalten mehr, sondern nur ein Teil, der auf einem niedrigeren Komplexitätsniveau analysiert wird. Dies war jahrzehntelang der Zugang zu diesem Problem, doch dieser Zugang ist unangebracht, da "das Bild vom Anderen" selbst zum sozialen Verhalten gehört. Zweckmäßiger ist, die "Richtigkeitsfrage" in den Hintergrund zu stellen und die Menschenbeurteilung als soziales Verhalten auf einem höheren Niveau der Komplexität zu betrachten und sie

dann mit anderen Aspekten sozialen Verhaltens, aber auch mit objektiv erhobenen Persönlichkeitsdaten und gegebenenfalls Leistungsmerkmalen, in Verbindung zu bringen. Dabei betrachtet man die Beurteilung der Mitmenschen als abhängig von:

1. der Eigenart des Beobachteten;
2. der Eigenart des Beobachters;
3. der Eigenart der Information, die der Beobachter über den Beobachteten bekommen hat.

Da Relationen zwischen diesen Punkten betrachtet werden, wird die Frage nach der "Richtigkeit" der Urteile abgelöst von der Frage: Wie spielt sich der Prozeß der Urteilsbildung ab?

In der Einleitung wurde bereits ausführlich dargelegt, daß Lehrer feste Vorstellungen über Eigenschaftszusammenhänge bei Schülern besitzen. Diese Eigenschaftszusammenhänge beschreibt HOFER (1969) durch die Faktoren Diszipliniertheit vs. Undiszipliniertheit, Zurückhaltung vs. Geltungsstreben, geistige Regsamkeit, soziale Aufgeschlossenheit vs. Verschlossenheit, Empfindsamkeit vs. Robustheit. HOFER (1969, S. 30-32) fand aus einer Primärliste von 191 Eigenschaftswörtern diejenigen Eigenschaften heraus, die Lehrer häufig zur Beurteilung von Schülern verwenden und die das Urteilsverhalten von Lehrern umfassend (repräsentativ) beschreiben. Außerdem schloß er homonyme und synonyme Eigenschaftsbegriffe aus. Sein Beurteilungsbogen zur aktuellen Beurteilung der Schüler umfaßt 25 Eigenschaften, die in gleichem Ausmaß positive wie negative Aspekte einschließen. Zwischen den polar angeordneten Eigenschaften dieses Bogens (Polaritätenprofil) bestehen sieben Möglichkeiten der Abstufung. Der Eigenschaftsbogen mit vorgegebener Instruktion ist dem Anhang beigefügt.

## 7.2. Ablauf der Datenerhebung

Die Erhebung der Daten, die in der Zeit von Januar bis März 1972 stattfand, wurde diesmal vom Verfasser selbst

durchgeführt.

Zunächst wurde den Schülern der 48-Item-Fragebogen zur Lehrerbeurteilung vorgelegt, im Anschluß daran der Lügenfragebogen. Der jeweilig zu beurteilende Klassenlehrer war während der Befragung nicht im Klassenraum. Um die Stichprobe homogener zu halten als in der Voruntersuchung, waren diesmal nur siebte, achte und neunte Klassen beteiligt. Außerdem ist in der Stichprobe ein zehntes Schuljahr enthalten. Die Befragungsdauer überschritt, selbst bei den Sonderschülern, nicht die 30-Minuten-Grenze. Verständnisschwierigkeiten traten, auch bei den Sonderschülern, kaum auf. Der Verfasser versuchte, durch informelle Gespräche solche Lehrer in den Kollegien zu finden, die nach Ansicht der jeweiligen Kollegen und Schulleiter unterschiedliche Unterrichtsstile praktizieren, um wieder (wie in der Voruntersuchung) eine Varianzvergrößerung der Schülerurteile über die Lehrer zu bewirken. Es gelang, auch solche Lehrer bzw. Lehrerinnen zu gewinnen, die der Befragungsaktion anfangs ablehnend und skeptisch gegenüberstanden.

An der Untersuchung nahmen Schulen aus Dortmund, Bochum, Düsseldorf und Osnabrück teil. Aus Gründen der Anonymität werden die beteiligten Schulen und Klassen nicht näher benannt. Tabelle 6 sind die an der Untersuchung beteiligten Jahrgänge und Schultypen zu entnehmen. Insgesamt waren an der Hauptuntersuchung 937 Schüler und Schülerinnen aus 40 Klassen beteiligt.

Tabelle 6

Zusammensetzung der Stichprobe der Hauptuntersuchung

| | Klasse (Jahrgang) | | | | |
|---|---|---|---|---|---|
| Schulart | 7 | 8 | 9 | 10 | Insgesamt |
| Hauptschule | 8 | 3 | 3 | 1 | 15 |
| Realschule | 3 | 4 | 3 | -- | 10 |
| Sonderschule für Lernbehinderte | 3 | 8 | 4 | -- | 15 |
| Insgesamt | 14 | 15 | 10 | 1 | 40 |

Das Alter der beteiligten Lehrer und Lehrerinnen schwankte zwischen 22 und 68 Jahren. Die Altersverteilung weist Tabelle 7 aus.

Tabelle 7
Altersverteilung der an der Untersuchung beteiligten Lehrer

| Alter | 20 - 30 | 31 - 40 | 41 - 50 | über 50 |
|---|---|---|---|---|
| Lehrer | 3 | 11 | 9 | 2 |
| Lehrerinnen | 4 | 7 | 4 | - |
| | 7 | 18 | 13 | 2 |

Im Anschluß an die Schülerbefragung wurden jedem beurteilten Klassenlehrer analog zur Voruntersuchung zwei Exemplare des Schülerfragebogens mit der Bitte vorgelegt, itemweise die durchschnittlich von ihren Schülern erhaltenen Beurteilungen zu schätzen (vermutetes Realurteil) und itemweise anzugeben, welche Idealbeurteilungen sie sich von ihren Schülern wünschen (Wunschurteil).

Zusätzlich erhielten die Lehrer noch die Polaritäten einschließlich Instruktion zur aktuellen Schülerbeurteilung mit der Bitte, jeden Schüler bzw. jede Schülerin ihrer jeweiligen Klasse anhand der vorgegebenen Polaritäten einzustufen.

Zusammenfassend erhielten wir also folgende Datensätze:

- Die Urteile aus den Fragebögen zum Interaktionsstil der Lehrer aus der Sicht der Schüler.

- Die Angaben aus dem Lügenfragebogen für jeden Schüler zur Erfassung von Simulationstendenzen der Schüler.

- Die itemweisen (durchschnittlichen) Schätzurteile der Lehrer (die vermuteten Realurteile)

- Die itemweisen Wunschurteile der Lehrer.

- Die Angaben der Lehrer aus dem Polaritätenprofil zur aktuellen Schülerbeurteilung für jeden ihrer Schüler.

- Daten zum Alter und Geschlecht der Lehrer, zum Schultyp und Jahrgang der Klassen, zum Alter und Geschlecht der Schüler.

## 7.3. Auswertung der Daten und Darstellung der Ergebnisse

### 7.3.1. Faktorenanalyse der Items

Die für den Schülerfragebogen erhaltenen Daten führten wir (analog zur Voruntersuchung) einer Hauptachsenanalyse mit anschließender Varimax-Rotation zu. Fragestellung 1 zufolge (vergl. Kapitel 6) versprachen wir uns Aufschluß darüber, ob auch für die Stichprobe der Hauptuntersuchung - bestehend aus Haupt-, Real- und Sonderschülern - eine dreifaktorielle Aufgliederung angemessen sei.

Nach dem Eigenwertdiagramm (s. Abb. 1) könnte man sich für drei zu extrahierende Faktoren entscheiden.

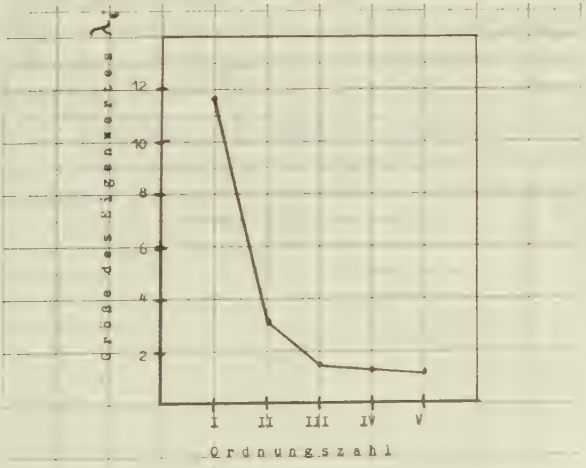



Abb. 1 : Eigenwertverlauf zur Faktoren-analyse der Items

Da jedoch bei der Dreierstruktur nur fünf Items Ladungshöhen über 0,3 auf dem dritten Faktor aufweisen (vergl. Tabelle 8a) - zudem haben nach Inspektion des Rechenprotokolls 36 Items Ladungen unter 0,20 auf dem dritten Faktor - entschieden wir uns, nur zwei Faktoren (Strenge und Unterstützung) zu interpretieren. Die Items ordneten wir gemäß ihrer Ladungshöhe jeweils der Strenge- bzw. Unterstützungsdimension zu, wobei Items mit etwa gleich hohen Ladungszahlen auf beiden Faktoren der Dimension zugewiesen wurden, auf der sie am höchsten laden. Tabelle 8b sind unter anderem die Faktorenladungen zu entnehmen.

Die neue Strengeskala enthält nun 25, die neue Unterstützungsskala 23 Items.

## 7.3.2. Aufgabenanalyse und Reliabilität für die Subskalen Strenge und Unterstützung

Analog zur Voruntersuchung erfolgten die Itemanalysen und Reliabilitätsschätzungen von diesmal nur zwei (faktorenanalytisch gewonnenen) Subskalen. Allerdings wurden zwei Itemanalysen gerechnet; die erste aufgrund der Gesamtstichprobe (N = 937); der zweiten Itemanalyse lag nur die Sonderschülerstichprobe zugrunde (N = 228). Tabelle 8b gibt die Trennschärfekoeffizienten $r_{it}$ und Schwierigkeitsindizes getrennt für beide Stichproben wieder.

Tab. 8a

*Ausschnitt aus dem Rechenprotokoll der R-Analyse:*
*Ladungszahlen der Items auf drei Faktoren*

Items

| FAKTOREN | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.34 | 0.38 | 0.08 | 0.20 | 0.27 | -0.27 | -0.50 | 0.25 | 0.42 | 0.21 | 0.09 | -0.12 | -0.38 |
| 2 | 0.37 | 0.33 | 0.14 | 0.09 | 0.52 | 0.17 | -0.27 | 0.54 | 0.37 | 0.17 | 0.33 | 0.44 | -0.11 |
| 3 | 0.01 | -0.14 | -0.57 | -0.26 | -0.06 | -0.37 | -0.06 | -0.09 | -0.18 | 0.58 | -0.24 | 0.06 | 0.07 |

| | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | -0.59 | 0.55 | 0.50 | 0.22 | 0.10 | 0.15 | -0.56 | 0.30 | 0.39 | 0.01 | -0.06 | 0.13 | -0.63 |
| 2 | -0.25 | 0.29 | 0.31 | 0.47 | 0.33 | 0.48 | -0.03 | 0.35 | 0.28 | 0.59 | 0.62 | 0.64 | -0.05 |
| 3 | 0.05 | -0.17 | 0.01 | 0.23 | -0.07 | 0.10 | 0.04 | -0.07 | -0.21 | 0.08 | -0.10 | -0.06 | -0.07 |

| | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.31 | 0.43 | 0.28 | 0.40 | -0.62 | 0.44 | -0.56 | 0.27 | -0.59 | 0.50 | 0.24 | 0.06 | -0.40 |
| 2 | 0.54 | 0.58 | 0.48 | 0.46 | -0.12 | 0.43 | 0.04 | 0.38 | -0.14 | 0.35 | 0.50 | 0.12 | -0.34 |
| 3 | 0.12 | 0.15 | 0.12 | 0.13 | -0.05 | -0.03 | -0.07 | 0.31 | -0.14 | 0.14 | -0.17 | 0.53 | -0.02 |

| | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.24 | 0.47 | -0.64 | 0.51 | 0.52 | -0.51 | 0.45 | -0.70 | -0.56 |
| 2 | 0.33 | 0.43 | -0.06 | 0.33 | 0.48 | -0.21 | 0.39 | -0.21 | -0.14 |
| 3 | -0.10 | 0.02 | -0.03 | 0.02 | 0.22 | 0.20 | 0.17 | -0.12 | -0.23 |

ROTATION MIT DELTA = 0.010

BETA = 0.881

EIGENWERTE

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 11.84 | 2.53 | 1.74 | 1.57 | 1.42 | 1.22 | 1.17 | 1.06 | 1.05 | 1.01 | 0.96 | 0.91 | 0.88 |

| 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.87 | 0.84 | 0.84 | 0.82 | 0.79 | 0.77 | 0.74 | 0.73 | 0.70 | 0.69 | 0.68 | 0.68 | 0.66 |

| 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.65 | 0.61 | 0.61 | 0.60 | 0.59 | 0.58 | 0.57 | 0.56 | 0.53 | 0.51 | 0.50 | 0.49 | 0.48 |

| 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 |
|---|---|---|---|---|---|---|---|---|
| 0.46 | 0.46 | 0.43 | 0.42 | 0.39 | 0.37 | 0.36 | 0.33 | 0.31 |

Tab. 8b

Schwierigkeiten, Trennschärfeindizes und Faktorenladungen der Items des Schülerfragebogens für die Hauptuntersuchung

I. STRENGE

| Item Nr. | Gesamtstichprobe p | $r_{it}$ | Ladungen I | II | Sonderschülerstichprobe p | $r_{it}$ | Ladungen I | II |
|---|---|---|---|---|---|---|---|---|
| 6 | .90 | .22 | -.36 | .25 | .92 | .08 | -.08 | .14 |
| 7 | .45 | .56 | -.50 | -.27 | .44 | .46 | -.46 | -.20 |
| 10 | .37 | .38 | .38 | .02 | .37 | .36 | .29 | .08 |
| 13 | .26 | .38 | -.34 | -.14 | .24 | .29 | -.29 | -.07 |
| 14 | .24 | .62 | -.55 | -.28 | .20 | .48 | -.51 | -.06 |
| 15 | .31 | .59 | .48 | .36 | .22 | .44 | .44 | .11 |
| 16 | .15 | .56 | .49 | .32 | .18 | .43 | .37 | .26 |
| 20 | .48 | .50 | -.52 | -.07 | .51 | .42 | -.51 | .15 |
| 26 | .51 | .58 | -.62 | -.07 | .55 | .45 | -.59 | .25 |
| 30 | .35 | .57 | .43 | .43 | .40 | .49 | .33 | .39 |
| 31 | .59 | .60 | -.61 | -.14 | .57 | .51 | -.57 | .03 |
| 33 | .37 | .50 | -.56 | .03 | .33 | .49 | -.53 | .05 |
| 34 | .30 | .48 | .36 | .30 | .25 | .45 | .36 | .27 |
| 35 | .46 | .60 | -.61 | -.14 | .44 | .54 | -.56 | -.08 |
| 36 | .36 | .61 | .52 | .33 | .34 | .47 | .38 | .21 |
| 38 | .49 | .25 | .22 | -.02 | .47 | .21 | .01 | .28 |
| 39 | .28 | .51 | -.39 | -.34 | .31 | .53 | -.40 | -.44 |
| 41 | .18 | .59 | .47 | .43 | .20 | .57 | .44 | .44 |
| 42 | .50 | .59 | -.62 | -.09 | .46 | .53 | -.59 | .06 |
| 43 | .18 | .60 | .50 | .35 | .24 | .56 | .43 | .43 |
| 44 | .18 | .68 | .57 | .44 | .20 | .57 | .45 | .41 |
| 45 | .39 | .52 | -.43 | -.29 | .30 | .46 | -.45 | -.16 |
| 46 | .35 | .60 | .49 | .35 | .35 | .48 | .31 | .42 |
| 47 | .29 | .71 | -.71 | -.21 | .50 | .57 | -.63 | -.04 |
| 48 | .42 | .60 | -.61 | -.11 | .38 | .47 | -.51 | -.06 |

II. UNTERSTÜTZUNG

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | .67 | .49 | .34 | .37 | .76 | .43 | .27 | .31 |
| 2 | .85 | .48 | .33 | .38 | .84 | .31 | .26 | .18 |
| 3 | .78 | .25 | -.09 | .29 | .87 | .15 | .00 | .05 |
| 4 | .65 | .27 | .12 | .17 | .73 | .36 | .25 | .20 |
| 5 | .76 | .57 | .26 | .53 | .85 | .52 | .26 | .44 |
| 8 | .74 | .59 | .22 | .56 | .81 | .43 | .19 | .40 |
| 9 | .63 | .56 | .35 | .43 | .67 | .49 | .46 | .27 |
| 11 | .72 | .38 | .01 | .39 | .81 | .30 | -.06 | .32 |
| 12 | .73 | .35 | -.10 | .42 | .79 | .35 | -.13 | .45 |
| 17 | .67 | .49 | .29 | .41 | .65 | .29 | .12 | .23 |
| 18 | .79 | .37 | .08 | .34 | .80 | .45 | .23 | .37 |
| 19 | .52 | .52 | .18 | .44 | .65 | .42 | .15 | .35 |
| 21 | .74 | .47 | .27 | .37 | .75 | .40 | .36 | .15 |
| 22 | .68 | .41 | .22 | .34 | .76 | .42 | .34 | .30 |
| 23 | .77 | .51 | .03 | .57 | .81 | .35 | -.10 | .50 |
| 24 | .65 | .52 | -.08 | .62 | .79 | .51 | -.02 | .55 |
| 25 | .67 | .61 | .12 | .64 | .75 | .45 | .12 | .48 |
| 27 | .73 | .60 | .35 | .31 | .69 | .54 | .24 | .48 |
| 28 | .79 | .68 | .47 | .55 | .79 | .57 | .31 | .48 |
| 29 | .67 | .52 | .31 | .45 | .61 | .39 | .04 | .39 |
| 32 | .79 | .58 | .42 | .45 | .83 | .48 | .47 | .29 |
| 37 | .88 | .54 | .19 | .54 | .89 | .41 | .15 | .43 |
| 40 | .90 | .40 | .21 | .36 | .83 | .29 | .17 | .23 |

Insgesamt fallen die Trennschärfekoeffizienten bei den Sonderschülern, verglichen mit der Gesamtstichprobe, erwartungsgemäß etwas ab, ebenfalls die Reliabilitätsschätzungen nach KUDER - RICHARDSON 21 (vergl. Tabelle 9). Trennschärfe- und Reliabilitätskoeffizienten können aber, bezogen auf die Sonderschülerstichprobe, als zufriedenstellend hoch bezeichnet werden (s. Fragestellung 2).

Tabelle 9
Reliabilitätskoeffizienten für die Gesamt- und Sonderschülerstichprobe

| | Gesamtstichprobe (N = 937) | Sonderschülerstichprobe (N = 228) |
|---|---|---|
| STRENGE | 0,90 | 0,84 |
| UNTERSTÜTZUNG | 0,85 | 0,77 |

Die gegenüber der Voruntersuchung höheren Reliabilitätskoeffizienten (0,90 für Strenge und 0,85 für Unterstützung) lassen sich vor allem auf die "Testverlängerung" für Strenge und Unterstützung zurückführen. Aufgrund der Voruntersuchung verteilen sich 44 Items auf drei, aufgrund der Hauptuntersuchung 48 Items auf zwei Dimensionen.

7.3.3. Korrelation zwischen den Beurteilungsmodi für Strenge und Unterstützung

Die Voruntersuchung erbrachte, daß, verglichen mit Unterstützung und Ausgleichendem Gewährenlassen, die Real-, Schätz- und Wunschurteile bezüglich Strenge relativ hoch miteinander korrelieren. Dieses Ergebnis, das wir dahingehend interpretierten, daß Lehrer das Ausmaß ihrer von den Schülern perzipierten Strenge besser einschätzen können

als das Ausmaß an Unterstützung und Ausgleichendem Gewährenlassen, soll (gemäß Fragestellung 3) noch einmal genauer überprüft werden.

Im Voruntersuchungsteil wurden die Zusammenhänge zwischen den Beurteilungsmodi (Real-, Schätz- und Wunschurteilen) dadurch gewonnen, daß die Summenscores der Beurteilungsmodi über die Anzahl der Lehrer korreliert wurden. Diesmal werden die Beurteilungsmodi für Strenge und Unterstützung klassenweise über die Itemscores korreliert. Die Korrelationskoeffizienten für Strenge basieren dabei auf 25 (weil 25 Items) und für Unterstützung auf 23 (weil 23 Items) Meßwertpaaren. Die Korrelation über die Itemscores gibt Aufschluß über die Ähnlichkeit zwischen den Beurteilungsmodi, d.h. inwieweit die auf die Realurteile bezogenen Rangpositionen der Items den Rangpositionen der Schätz- und Wunschurteile einerseits und die auf die Schätzurteile bezogenen Rangpositionen der Items den Rangpositionen der Wunschurteile andererseits entsprechen. Wie aus Tabelle 10 hervorgeht, kovariieren die Beurteilungsmodi für Strenge im Durchschnitt höher als für Unterstützung. Das wird auch an der hohen Anzahl der signifikanten Korrelationen zwischen den Beurteilungsmodi bezüglich Strenge deutlich. Auffällig ist die relativ hohe durchschnittliche Korrelation zwischen den Schätz- und Wunschurteilen für Strenge und insbesondere für Unterstützung. Interpretativ sei hier vorweggenommen, daß sich bei den untersuchten Lehrern die Schätz- und Wunschurteile stärker ähneln als daß sich die Schätz- und Wunschurteile an den Realurteilen ausrichten. Dies gilt insbesondere für Unterstützung.

### 7.3.4. Kritik an den faktorenanalytisch gewonnenen Subskalen Strenge und Unterstützung

Wir hatten die Faktorenanalyse als Hilfsmittel zur Bestimmung der Dimensionalität des Itempools herangezogen. Die Fak-

Tab. 10

Korrelationen zwischen den Beurteilungsmodi für Strenge und Unterstützung

| Lehrer Nr. | STRENGE | | | UNTERSTÜTZUNG | | |
|---|---|---|---|---|---|---|
| | $r_{real/schätz}$ | $r_{real/wunsch}$ | $r_{schätz/wunsch}$ | $r_{real/schätz}$ | $r_{real/wunsch}$ | $r_{schätz/wunsch}$ |
| 1 | .71 | .56 | .54 | .23 | .58 | .70 |
| 2 | .58 | .58 | .97 | .30 | .30 | 1.00 |
| 3 | .33 | .38 | .49 | .74 | .27 | .59 |
| 4 | .53 | .07 | .08 | .15 | .02 | .28 |
| 5 | .45 | .60 | .70 | .43 | .20 | .59 |
| 6 | .34 | .48 | .59 | .49 | -.03 | .21 |
| 7 | .61 | .54 | .58 | -.11 | -.23 | .10 |
| 8 | .65 | .62 | .33 | .36 | .16 | .51 |
| 9 | .49 | .49 | .75 | .30 | .18 | .61 |
| 10 | .70 | .72 | .86 | .30 | .30 | .69 |
| 11 | .42 | .60 | .56 | .18 | .23 | .79 |
| 12 | .70 | .50 | .46 | .17 | .01 | .18 |
| 13 | .67 | .78 | .67 | .13 | .11 | .69 |
| 14 | .68 | .58 | .86 | .17 | -.07 | .74 |
| 15 | .62 | .62 | .66 | .04 | -.04 | .12 |
| 16 | .66 | .42 | .58 | .42 | .26 | -.18 |
| 17 | .61 | .55 | .67 | .65 | .01 | .25 |
| 18 | .60 | .62 | .67 | .65 | .32 | .25 |
| 19 | .62 | .49 | .74 | .68 | .34 | .55 |
| 20 | .62 | .76 | .64 | .31 | .33 | .19 |
| 21 | .73 | .63 | .58 | .33 | .20 | .32 |
| 22 | .82 | .83 | .87 | .35 | .39 | .84 |
| 23 | .63 | .55 | .86 | .59 | .20 | .37 |
| 24 | .73 | .76 | .67 | .61 | .28 | .62 |
| 25 | .74 | .54 | .53 | .47 | -.04 | .54 |
| 26 | .67 | .71 | .80 | .41 | .40 | .25 |
| 27 | .38 | .41 | .91 | .04 | .11 | .88 |
| 28 | .50 | .56 | .65 | .36 | .41 | .34 |
| 29 | .27 | .40 | .45 | .59 | .22 | .48 |
| 30 | .75 | .74 | .81 | .17 | .33 | .38 |
| 31 | .48 | .70 | .20 | .41 | .30 | .31 |
| 32 | .81 | .78 | .95 | .54 | .54 | 1.00 |
| 33 | .57 | .72 | .59 | .34 | .45 | .73 |
| 34 | .63 | .63 | .97 | .51 | .51 | 1.00 |
| 35 | .54 | .62 | .58 | .24 | .20 | .44 |
| 36 | .75 | -.01 | -.16 | .51 | .01 | .45 |
| 37 | .76 | .57 | .75 | .25 | .07 | .75 |
| 38 | .65 | .65 | .66 | .80 | .25 | .57 |
| 39 | .57 | .68 | .73 | .59 | .45 | .63 |
| 40 | .80 | .70 | .45 | .74 | -.15 | .20 |
| $\bar{r}$ | .63 | .60 | .69 | .41 | .22 | .63 |
| Anzahl signifikanter Korrelationen | 37 | 37 | 36 | 19 | 6 | 24 |

torenanalyse erfüllte dabei eine vorläufige Ordnungsfunktion, der gesamte Itempool wurde in zwei homogene Itemgruppen aufgegliedert.

In der klassischen Testtheorie ist Itemhomogenität, die man auch faktorielle Homogenität nennt, durch hohe Iteminterkorrelationen bedingt. Es läßt sich zeigen, daß die sogenannte faktorielle Homogenität, insbesondere wenn dichotome, oder wie in unserem Fall dichotome Items nahekommende vierstufige Antwortskalen verwendet werden, bei ähnlichen Schwierigkeitsindizes erfüllt ist (HORST, 1966, S. 193-209). Während Items gleicher oder ähnlicher Schwierigkeit hoch miteinander korrelieren, führen Items unterschiedlicher Schwierigkeit zu niedrigen Korrelationskoeffizienten. Eine anschließende Faktorenanalyse bringt dann auf den extrahierten Faktoren lediglich die nach ihrer Schwierigkeit geordneten Items.

Neben der faktoriellen Homogenität gibt es den strengen (kumulativen) Homogenitätsbegriff GUTTMANNS (1950, 1955). Hier wird ein eindimensionales Kontinuum postuliert, auf dem Vpn. gemäß ihrem Ausprägungsgrad der zu messenden Eigenschaft durch Beantwortung der nach ihrer Schwierigkeit geordneten Items lokalisierbar sind.

Beim GUTTMANN - Modell ist also jeder "Kontakt" zwischen einer Vp. und einem Item durch den Personenparameter $\xi$ und den Itemparameter $\sigma$ eindeutig determiniert, d.h., daß sich aus $\xi$ und $\sigma$ zwingend ein bestimmtes Resultat ergibt. GUTTMANNS Modell ist somit deterministisch und stellt Anforderungen an die empirischen Daten, die selten erfüllt sind (vergl. FISCHER, 1968, S. 73-74).

würde man eine perfekte GUTTMANN-Skala faktorisieren, so erhielten wir durchaus nicht einen Faktor, sondern soviele Faktoren wie unterschiedlich schwierige Items (vergl. HORST, 1966, S. 161).

Betrachten wir die durch die Faktorenanalyse erhaltenen zwei homogenen Itemgruppen (die Strenge- und Unterstützungsitems), so fallen die großen numerischen Unterschiede der Schwierigkeitsindizes zwischen den Strenge- und Unterstützungsitems auf, es handelt sich geradezu um zwei verschiedene Schwierigkeitsclusters ($\bar{P}_{Strenge} = 0,37$; $\bar{P}_{Unterstützung} = 0,73$). Das Auffinden der beiden Faktoren Strenge und Unterstützung ist also auf den Umstand zweier verschiedener Schwierigkeitsclusters der Items, also auf die unterschiedliche Beantwortungshäufigkeit von Strenge- und Unterstützungsitems durch die Schülerstichprobe zurückzuführen. Strenge- und Unterstützungsitems mit homogenen (d.d. nicht unterschiedlichen) Schwierigkeiten würden sich vermutlich auf einer gemeinsamen (bipolaren) Dimension lokalisieren lassen, wodurch wir eine gemischte Skala aus Strenge- und Unterstützungsitems erhalten hätten.

Die bisherigen Ausführungen zur Kritik an unseren faktorenanalytisch gewonnene Subskalen mögen verdeutlicht haben, warum es problematisch ist, von homogenen Itemgruppen auf unabhängigen Dimensionen zu sprechen.

Würden wir die beiden homogenen Itemgruppen gesondert einer Faktorenanalyse unterziehen, so müßten wir gleichzeitig auf eine allgemeine Kritik an der Faktorenanalyse hinweisen. Überlegungen hierzu sollen an dieser Stelle nicht vertieft werden, es sei nur daran erinnert, daß die Zahl der Faktoren und die Art der Faktorenstruktur von den Eigenschaften der Referenzpopulation abhängen. Außerdem bedingen Zahl und Art der Variablen bzw. Items die Faktorenstruktur und -anzahl mit; wenn homogene Merkmale gesondert analysiert werden, sind weitere Aufschlüsselungen möglich. Es fehlen Kriterien zur Vereinbarung eines festgelegten Komplexitätsniveaus.

Stellen wir die Kritik an den faktorenanalytisch gewonnenen Subskalen in Rechnung, so sollten wir nach Möglichkeiten einer verbesserten Skalenkonstruktion Ausschau halten.

## 7.3.5. Skalenkonstruktion anhand interner und externer Selektionskriterien

### 7.3.5.1. Vorbemerkungen

Da wir wegen der diskutierten methodologischen Einwände auf die Faktorenanalyse als Hilfsmittel zur Skalengewinnung verzichten, stellt sich an dieser Stelle das Problem, ein geeigneteres Skalierungsmodell zur Skalenkonstruktion und Itemselektion zu wählen.

Die sogenannten klassischen Skalierungstechniken nach THURSTONE und LIKERT scheiden wegen unklarer Modelleigenschaften aus. Das GUTTMANN-Modell ist, wie bereits diskutiert, deterministisch und stellt Anforderungen an die empirischen Daten, die selten erfüllt sind. Demgegenüber wäre die Entscheidung für ein stochastisches Modell angezeigt, zumal die Vorzüge von stochastischen Modellen darin liegen, daß sie den empirischen Daten weiten Spielraum lassen, da Zufallsprozesse essentielle Modellbestandteile sind (FISCHER, 1968, S. 74) und speziell mit der RASCH'schen probabilistischen Testtheorie bereits praktische Erfahrungen im Rahmen der Erziehungsstilforschung vorliegen (s. STAPF et al., 1971, S. 90 f).

Greifen wir von den stochastischen Maßmodellen beispielsweise das probabilistische Modell von RASCH (1960, 1966a, 1966b; STENE, 1968) heraus, so bieten sich folgende Vorteile an:

- Das Modell akzeptiert bei der iterativen Itemselektion nur homogene Items mit monotonen Itemcharakteristiken, es gestattet also, eindimensionale Messungen vorzunehmen.

- Der Vorzug der spezifischen Objektivität ist gegeben. Es ist das Ziel des Modells, Itemparameter zu definieren,

die sich unabhängig von den Personenparametern schätzen lassen und umgekehrt. Die Möglichkeit, eine simultane Metrik für Vpn und Tests zu erstellen, ist neu. Außerdem erhält man die Item- und Personenparameter auf Verhältnisskalenniveau.

- Das RASCH'sche Modell erlaubt, populationsunabhängige Messungen vorzunehmen, was unseren Bemühungen entgegenkäme, für Haupt-, Real- und insbesondere Sonderschüler gleichermaßen gut geeignete Items zu finden.

Allerdings fehlen probabilistischen Testmodellen bis heute die modellgerechten Ansätze zu Validitätsfragestellungen. Man ist bezüglich der Validitätsüberprüfungen wieder auf die Techniken der klassischen Test- und Meßmodelle angewiesen. Deshalb entschieden wir uns für klassische Techniken zur Skalenkonstruktion und Itemselektion. Neben internen Selektionskriterien (Trennschärfe und Schwierigkeit) werden zwei externe Selektionskriterien, von denen in den folgenden Kapiteln die Rede sein wird, zur Itemselektion herangezogen. Schließlich kommt noch eine varianzanalytische Überprüfung der Itembrauchbarkeit zum Tragen, die später zu diskutieren sein wird.

### 7.3.5.2. Itemauswahl aufgrund von Expertenrating

Rufen wir uns noch einmal ins Gedächtnis, daß es das Hauptanliegen dieser Untersuchung ist, für Klima und Lernatmosphäre in Klassen bedeutsame Lehrermerkmale zu finden.

Aus dem komplexen Forschungsgegenstand Lehrer-Schüler-Interaktion hatten wir einen wichtigen Teilbereich, nämlich solche lernpsychologisch spezifizierten Verhaltensmerkmale von Lehrern herausgegriffen, die die Bekräftigungs-

wirkung des Lehrers auf einer molaren Verhaltensebene charakterisieren sollen (vergl. Kapitel 2.5.). Hierbei ist unser Vorgehen dem bei der Entwicklung psychodiagnostischer Verfahren analog: Die individuellen Eigenarten von Lehrern bezüglich bestimmter Merkmale (ihres Bekräftigungsverhaltens) werden festgestellt. Aufgrund eines relativ kleinen Verhaltensausschnittes, d.h. einer Stichprobe von Verhaltensmerkmalen bzw. verhaltenskorrelierenden Merkmalen wird auf das Gesamtverhalten eines Lehrers geschlossen (soweit es sein lernpsychologisch bedeutsames Bekräftigungsverhalten betrifft). Die Güte unseres (diagnostischen) Beobachtungsverfahrens ist somit ein Stichprobenproblem.

In den bisherigen Ausführungen zur Itembrauchbarkeit (s. Tabelle 7 im Kapitel 8.3.2.) hatten wir die Trennschärfekoeffizienten und Schwierigkeitsindizes der Items im Hinblick auf die beiden Subskalen Strenge und Unterstützung als interne Selektionskriterien herausgestellt. Ein Expertenrating soll nun als externes Selektionsverfahren Aufschlüsse darüber erbringen, welche Items des Itempools sich am besten zur Charakterisierung a n g e b r a c h t e n Interaktionsverhaltens des Lehrers, bezogen auf Lernatmosphäre und Klassenklima, eignen (s. Fragestellung 4).

Wenn wir Experten danach fragen, welche Items in welchem Ausmaß angemessenes Interaktionsverhalten des Lehrers kennzeichnen, reduzieren wir das zuvor im Kapitel 8.3.4. kritisierte zweidimensionale Konzept Strenge und Unterstützung auf e i n e Dimension, die wir in Anlehnung an unsere lernpsychologische Terminologie "Bekräftigungswirkung des Lehrers" nennen wollen und die als eine bipolare Dimension unterstellt wird.

Die eindimensionalen Skalierungsverfahren werden nach TORGERSON (1958, S. 55 - 58) in quantitativ und variability judgment methods eingeteilt. Bei den quantitativ jugdment methods, zu denen unter anderem die subjektiven Schätzmethoden, die Fraktionierungsmethoden und Gleichteilungsmethoden gehören, wirken sich Unterschiede zwischen den Ratern als störend aus. Bei diesen Methoden läuft die Skalierung dann besonders gut, wenn die Rater (unsere Experten) alle möglichst einheitlich urteilen, in diesem Falle lassen sich sehr viel genauer Stimulus-Mittelwerte errechnen. Da wir aber nicht erwarten können, daß die Experten bezüglich der zu skalierenden Stimuli (Items) einheitliche Urteile abgeben und wir größere Abweichungen zwischen den Experten nicht als Meßfehler tolerieren können, kommt für unser Expertenrating nur ein Verfahren der variability judgment methods infrage. Diese Methoden lassen nicht nur Urteilsunterschiede zu, sondern Urteilsvariabilität zwischen Ratern ist hier eine notwendige Voraussetzung und stellt eine Möglichkeit dar, als Endergebnis die Stimuli (Items) auf einer Intervallskala anzuordnen.

Da von den variability judgment methods die Methode des Paarvergleichs zu aufwendig ist, entschieden wir uns für das Skalierungsverfahren der sukzessiven Intervalle in Verbindung mit dem Gesetz vom kategorialen Urteil. Die theoretisch-mathematische Begründung, eine detaillierte Beschreibung sowie eine kritische Würdigung dieses Skalierungsverfahrens sind TORGERSON (1958, S. 205-246) zu entnehmen. Hier werden nur die für unser praktisches Problem notwendigen Schritte der Erhebung und rechnerischen Auswertung dargelegt.

Die 48 Items des Schülerfragebogens wurden einzeln auf Kärtchen geschrieben, die die Experten in eine Rangreihe von fünf Haufen zu ordnen hatten (Sortiermethode). Die Instruktion für die Experten lautete:

"Ich lege Ihnen 48 Kärtchen vor, auf denen Verhaltensweisen von Lehrern stehen. Bitte verteilen Sie die Karten auf fünf Haufen, wobei jeder Haufen einen Ausprägungsgrad darstellt. Ordnen Sie die Karten danach, in welchem Ausmaß (bzw. wie intensiv) jede Verhaltensweise zu einem guten Klassenklima und zu einer günstigen Lernatmosphäre in der Schulklasse beiträgt. Auf Haufen fünf sollen nur Verhaltensweisen fallen, die nach Ihrer Meinung maximal gut zur Förderung des Klassenklimas geeignet sind, Haufen eins soll solche Verhaltensweisen erhalten, die nicht zu einem guten Klassenklima beitragen. Die restlichen drei Haufen stellen Abstufungen zwischen diesen Extremen dar."

Bei den Experten (N = 33) handelte es sich um Hochschulpersonal (Pädagogen und Psychologen).

Die analytische Auswertung der Raterurteile erfolgte nach TORGERSON (1960, S. 211-213) in sechs Schritten, wobei jeder Schritt durch eine Matrix gekennzeichnet ist:

1. Schritt: F - Matrix:
In diese Matrix werden die Ergebnisse der Raterurteile eingetragen. Die Zeilen dieser Matrix geben die Häufigkeit an, wie oft Item j in die kategorie g eingestuft wurde. Es gibt eine Kategoriengrenze weniger als es Kategorien gibt, da die erste und letzte Kategorie offenstehen. Weil die Anzahl der Kategoriengrenzen gleich m ist, gibt es also m+1 Kategorien.

2. Schritt: $\int$ -Matrix:
Die $\int$ - Matrix enthält die von links aufkumulierten Häufigkeiten der Matrix F.

3. Schritt: P - Matrix:
Die P - Matrix enthält die auf 1 relativierten Werte der $\int$ - Matrix (Werte der $\int$ - Matrix durch N dividiert). Die letzte Spalte (m+1) entfällt, da in dieser Spalte überall die Zahl 1 steht.

4. Schritt: X - Matrix:
Die X - Matrix ergibt sich aus der P - Matrix durch z - Transformation. Folgendes ist zu beachten: Tritt in der X - Matrix 0 oder 1 auf, so erhält man durch Transformation für $z = \pm\infty$. Da man mit unendlichen Werten nicht weiterrechnen kann, läßt man diese Felder leer; sie sind durch einen Querstrich gekennzeichnet.

5. Schritt: X - Matrix (Differenzenmatrix):
Durch die Differenzenmatrix werden die Kategoriengrenzen ermittelt. Hierbei ist die Differenz nebeneinanderstehender z - Werte gleich der Differenz zwischen den entsprechenden Kategoriengrenzen. Wegen der Meßfehler erhält man jedoch nur Schätzwerte der Differenz tg+1 - tg. Die beste Schätzung der Differenz ist der Mittelwert über alle Stimuli j. Wir haben somit eine Möglichkeit, den Abstand zwischen aufeinanderfolgenden Kategoriengrenzen zu schätzen. Zu den Skalenwerten gelangt man, indem man zunächst die erste Kategoriengrenze $t_1 = 0$ setzt. Dann liegen die anderen Kategoriengrenzen tg ebenfalls fest, da die Differenzen bekannt sind.

6. Schritt: (( tg - zjg )) - Matrix:
Die Skalenwerte Sj der Stimuli erhält man durch Bildung der Differenzen tg - zjg für jede Kategorie; die beste Schätzung für Sj ist der Mittelwert der Differenzen.
- Der Beschreibung der Rechenschritte folgt nun die Verrechnung des Zahlenmaterials des Expertenratings. -

Tab. 11

Auswertung der Daten des Expertenratings

| Item Nr. | P - Matrix Kategorien 1 | 2 | 3 | 4 | 5 | $\hat{P}$ - Matrix Kategorien 1 | 2 | 3 | 4 | 5 | P - Matrix Kategorien 1 | 2 | 3 | 4 | X - Matrix Kategorien 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | - | 1 | 2 | 4 | 26 | - | 1 | 3 | 7 | 33 | - | 0.03 | 0.09 | 0.21 | - | -1.88 | -1.34 | -0.80 |
| 2 | - | - | 1 | 11 | 21 | - | - | 1 | 12 | 33 | - | - | 0.03 | 0.36 | - | - | -1.88 | -0.35 |
| 3 | 2 | 13 | 14 | 4 | - | - | 2 | 15 | 29 | 33 | - | 0.06 | 0.46 | 0.88 | - | -1.55 | -0.10 | 1.17 |
| 4 | - | 1 | 10 | 13 | 9 | - | 1 | 11 | 24 | 33 | - | 0.03 | 0.33 | 0.73 | - | -1.88 | -0.44 | 0.61 |
| 5 | - | - | - | 13 | 20 | - | - | - | 13 | 33 | - | - | 0.39 | 1.00 | - | - | -0.28 | - |
| 6 | 8 | 14 | 9 | 2 | - | 8 | 22 | 31 | 33 | 33 | 0.24 | 0.67 | 0.94 | 1.00 | -0.70 | 0.44 | 1.55 | - |
| 7 | 20 | 7 | 6 | - | - | 20 | 27 | 33 | 33 | 33 | 0.61 | 0.82 | 1.00 | 1.00 | 0.28 | 1.17 | - | - |
| 8 | - | - | 1 | 14 | 18 | - | - | 1 | 15 | 33 | - | - | 0.03 | 0.46 | - | - | -1.88 | -0.10 |
| 9 | 26 | 5 | 2 | - | - | 26 | 31 | 33 | 33 | 33 | 0.79 | 0.94 | 1.00 | 1.00 | 0.81 | 1.55 | - | - |
| 10 | 2 | 9 | 14 | 7 | 1 | 2 | 11 | 25 | 32 | 33 | 0.06 | 0.33 | 0.76 | 1.00 | -1.55 | -0.44 | 0.71 | - |
| 11 | 3 | 9 | 11 | 7 | 3 | 3 | 12 | 23 | 30 | 33 | 0.09 | 0.36 | 0.70 | 0.91 | -1.34 | -0.35 | 0.52 | 1.34 |
| 12 | 15 | 9 | 5 | 3 | 1 | 15 | 24 | 29 | 32 | 33 | 0.46 | 0.73 | 0.88 | 0.97 | -0.10 | 0.61 | 1.17 | 1.88 |
| 13 | 29 | 4 | - | - | - | 29 | 33 | 33 | 33 | 33 | 0.88 | 1.00 | 1.00 | 1.00 | 1.17 | - | - | - |
| 14 | 30 | 3 | - | - | - | 30 | 33 | 33 | 33 | 33 | 0.91 | 1.00 | 1.00 | 1.00 | 1.34 | - | - | - |
| 15 | 17 | 11 | 4 | 1 | - | 17 | 28 | 32 | 33 | 33 | 0.52 | 0.85 | 0.97 | 1.00 | 0.05 | 1.04 | 1.88 | - |
| 16 | - | - | 3 | 7 | 23 | - | - | 3 | 10 | 33 | - | - | 0.09 | 0.30 | - | - | -1.34 | -0.52 |
| 17 | - | 2 | 6 | 12 | 13 | - | 2 | 8 | 20 | 33 | - | 0.06 | 0.24 | 0.61 | - | -1.55 | -0.70 | 0.28 |
| 18 | - | 1 | 1 | 11 | 20 | - | 1 | 2 | 13 | 33 | - | 0.03 | 0.06 | 0.39 | - | -1.88 | -1.55 | -0.28 |
| 19 | - | - | 1 | 9 | 23 | - | - | 1 | 10 | 33 | - | - | 0.03 | 0.30 | - | - | -1.88 | -0.52 |
| 20 | 17 | 13 | 3 | - | - | 17 | 30 | 33 | 33 | 33 | 0.51 | 0.91 | 1.00 | 1.00 | 0.03 | 1.34 | - | - |
| 21 | 16 | 11 | 5 | - | 1 | 16 | 27 | 32 | 32 | 33 | 0,49 | 0.82 | 0.97 | 1.00 | -0.03 | 1.17 | 1.88 | - |
| 22 | - | - | 6 | 5 | 22 | - | - | 6 | 11 | 33 | - | - | 0.18 | 0.33 | - | - | -0.92 | -0.44 |
| 23 | - | - | 2 | 12 | 19 | - | - | 2 | 14 | 33 | - | - | 0.06 | 0.42 | - | - | -1.55 | -0.20 |
| 24 | - | 4 | 10 | 14 | 5 | - | 4 | 14 | 28 | 33 | - | 0.12 | 0.42 | 0.85 | - | -1.17 | -0.20 | 1.04 |
| 25 | - | - | 2 | 8 | 23 | - | - | 2 | 10 | 33 | - | - | 0.06 | 0.30 | - | - | -1.55 | -0.52 |
| 26 | 21 | 11 | 1 | - | - | 21 | 32 | 33 | 33 | 33 | 0.64 | 0.97 | 1.00 | 1.00 | 0.36 | 1.88 | - | - |
| 27 | - | - | 1 | 6 | 26 | - | - | 1 | 7 | 33 | - | - | 0.03 | 0.21 | - | - | -1.88 | -0.80 |
| 28 | - | 1 | 1 | 4 | 27 | - | 1 | 2 | 6 | 33 | - | 0.03 | 0.06 | 0.18 | - | -1.88 | -1.55 | -0.92 |
| 29 | 1 | 2 | 5 | 16 | 5 | 1 | 3 | 12 | 28 | 33 | 0.03 | 0.09 | 0.36 | 0.85 | -1.88 | -1.34 | -0.35 | 1.04 |
| 30 | 1 | 2 | 2 | 7 | 21 | 1 | 3 | 5 | 12 | 33 | 0.03 | 0.09 | 0.15 | 0.36 | -1.88 | -1.34 | -1.04 | -0.36 |
| 31 | 18 | 11 | 4 | - | - | 18 | 29 | 33 | 33 | 33 | 0.55 | 0.88 | 1.00 | 1.00 | 0.13 | 1.17 | - | - |
| 32 | - | 3 | 4 | 6 | 20 | - | 3 | 7 | 13 | 33 | - | 0.09 | 0.21 | 1.00 | - | 1.34 | -0.80 | - |
| 33 | 25 | 5 | 1 | 2 | - | 25 | 30 | 32 | 33 | 33 | 0.76 | 0.91 | 0.94 | 1.00 | 0.71 | 1.34 | 1.55 | - |
| 34 | - | 1 | 7 | 17 | 8 | - | 1 | 8 | 25 | 33 | - | 0.03 | 0.32 | 0.76 | - | -1.88 | -0.47 | 0.71 |
| 35 | 14 | 13 | 5 | 1 | - | 14 | 27 | 32 | 33 | 33 | 0.42 | 0.82 | 0.97 | 1.00 | -0.20 | 1.17 | 1.88 | - |
| 36 | - | 3 | 3 | 12 | 15 | - | 3 | 6 | 18 | 33 | - | 0.09 | 0.18 | 0.55 | - | -1.34 | -0.92 | 0.13 |
| 37 | - | - | 3 | 12 | 18 | - | - | 3 | 15 | 33 | - | - | 0.09 | 0.46 | - | - | -1.34 | -0.10 |
| 38 | 6 | 10 | 14 | 2 | 1 | 6 | 16 | 30 | 32 | 33 | 0.18 | 0.49 | 0.91 | 0.97 | -0.92 | -0.02 | 1.34 | 1.88 |
| 39 | 2 | 8 | 8 | 11 | 5 | 2 | 10 | 18 | 29 | 33 | 0.06 | 0.30 | 0.55 | 0.88 | -1.55 | -0.52 | 0.13 | 1.17 |
| 40 | - | 2 | 12 | 12 | 7 | - | 2 | 14 | 26 | 33 | - | 0.06 | 0.42 | 0.79 | - | -1.55 | -0.20 | 0.81 |
| 41 | 1 | 5 | - | 10 | 17 | 1 | 6 | 6 | 16 | 33 | 0.03 | 0.18 | 0.18 | 0.49 | -1.88 | -0.92 | -0.92 | -0.02 |
| 42 | 21 | 9 | 3 | - | - | 21 | 30 | 33 | 33 | 33 | 0.64 | 0.91 | 1.00 | 1.00 | 0.36 | 1.34 | - | - |
| 43 | 1 | 6 | 7 | 7 | 12 | 1 | 7 | 14 | 21 | 33 | 0.03 | 0.21 | 0.42 | 0.64 | -1.88 | -0.81 | -0.20 | 0.36 |
| 44 | - | - | - | 4 | 29 | - | - | - | 4 | 33 | - | - | - | 0.12 | - | - | - | -1.17 |
| 45 | 11 | 18 | 4 | - | - | 11 | 29 | 33 | 33 | 33 | 0.33 | 0.88 | 1.00 | 1.00 | -0.44 | 1.17 | - | - |
| 46 | - | 1 | 8 | 11 | 13 | - | 1 | 9 | 20 | 33 | - | 0.03 | 0.27 | 0.61 | - | -1.88 | -0.61 | 0.28 |
| 47 | 17 | 13 | 3 | - | - | 17 | 30 | 33 | 33 | 33 | 0.52 | 0.91 | 1.00 | 1.00 | 0.05 | 1.34 | - | - |
| 48 | 10 | 15 | 7 | 1 | - | 10 | 25 | 32 | 33 | 33 | 0.30 | 0.76 | 0.97 | 1.00 | -0.52 | 0.71 | 1.88 | - |

Fortsetzung von Tab. 11

| Item Nr. | ΔX - Matrix | | | (( $t_g - z_{jg}$ )) - Matrix | | | | |
|---|---|---|---|---|---|---|---|---|
| | 2 - 1 | 3 - 2 | 4 - 3 | $t_1 - z_{j1}$ | $t_2 - z_{j2}$ | $t_3 - z_{j3}$ | $t_4 - z_{j4}$ | Skalenwerte |
| 1 | — | 0.54 | 0.54 | - | 2.91 | 3.19 | 3.59 | 3.23 |
| 2 | - | - | 1.53 | | - | 3.73 | 3.14 | 3.43 |
| 3 | - | 1.45 | 1.27 | - | 2.58 | 1.95 | 1.62 | 2.05 |
| 4 | - | 1.44 | 1.05 | - | 2.91 | 2.29 | 2.18 | 2.46 |
| 5 | - | - | - | - | - | 2.13 | - | 2.13 |
| 6 | 1.14 | 1.11 | - | 0.70 | 0.59 | 0.30 | - | 0.53 |
| 7 | 0.89 | - | - | -0.28 | -0.14 | - | - | -0.21 |
| 8 | - | - | 1.78 | - | - | 3.73 | 2.89 | 3.31 |
| 9 | 0.74 | - | - | -0.81 | -0.52 | - | - | -0.55 |
| 10 | 1.11 | 1.15 | - | 1.55 | 1.47 | 1.14 | - | -0.66 |
| 11 | 0.99 | 0.87 | 0.82 | 1.34 | 1.38 | 1.33 | 1.45 | 1.37 |
| 12 | 0.71 | 0.56 | 0.71 | 0.10 | 0.42 | 0.68 | 0.91 | 0.52 |
| 13 | - | - | - | -1.17 | - | - | - | -1.17 |
| 14 | - | - | - | -1.34 | - | - | - | -1.34 |
| 15 | 0.99 | 0.84 | - | -0.05 | -0.01 | -0.03 | - | -0.03 |
| 16 | - | - | 0.82 | - | - | 3.19 | 3.31 | 3.25 |
| 17 | - | 0.85 | 0.98 | - | 0.52 | 2.55 | 2.51 | 1.86 |
| 18 | - | 0.33 | 0.27 | - | 0.85 | 3.40 | 3.07 | 2.44 |
| 19 | - | - | 0.36 | - | - | 3.73 | 3.31 | 3.52 |
| 20 | 1.31 | - | - | -0.03 | -0.31 | - | - | -0.17 |
| 21 | 1.20 | 0.71 | - | 0.03 | -0.14 | -0.03 | - | -0.04 |
| 22 | - | - | 0.48 | - | - | 2.77 | 3.23 | 3.00 |
| 23 | - | - | 1.35 | - | - | 3.40 | 2.99 | 3.19 |
| 24 | - | 0.97 | 1.24 | - | 2.20 | 2.05 | 1.75 | 2.00 |
| 25 | - | - | 1.03 | - | - | 3.40 | 3.31 | 3.35 |
| 26 | 1.52 | - | - | -0.36 | -0.85 | - | - | 0.60 |
| 27 | - | - | 1.08 | - | - | 3.73 | 3.59 | 3.66 |
| 28 | - | 0.33 | 0.63 | - | 2.91 | 3.40 | 3.71 | 3.34 |
| 29 | 0.54 | 0.99 | 1.39 | 1.88 | 2.37 | 2.20 | 1.75 | 2.05 |
| 30 | 0.54 | 0.30 | 0.68 | 1.88 | 2.37 | 2.89 | 3.05 | 2.54 |
| 31 | 1.04 | - | - | -0.13 | -0.14 | - | - | 0.13 |
| 32 | - | 0.54 | - | - | 2.37 | 2.65 | - | 2.51 |
| 33 | 0.63 | 0.21 | - | -0.71 | -0.31 | 0.30 | - | -0.36 |
| 34 | - | 1.41 | 1.18 | - | 2.91 | 2.32 | 2.08 | 2.43 |
| 35 | 1.37 | 0.71 | - | 0.20 | -0.14 | -0.03 | - | 0.10 |
| 36 | — | 0.42 | 1.05 | - | 2.37 | 2.77 | 2.66 | 2.60 |
| 37 | - | - | 1.24 | - | - | 3.19 | 2.89 | 3.04 |
| 38 | 0.90 | 1.36 | 0.54 | 0.92 | 1.05 | 0.51 | 0.91 | 0.84 |
| 39 | 1.03 | 0.65 | 1.04 | 1.55 | 1.55 | 1.72 | 1.62 | 1.61 |
| 40 | - | 1.35 | 1.01 | - | 2.58 | 2.05 | 1.98 | 2.20 |
| 41 | 0.96 | 0.00 | 0.90 | 1.88 | 1.95 | 2.77 | 2.81 | 2.35 |
| 42 | 0.98 | - | - | -0.36 | -0.31 | - | - | -0.33 |
| 43 | 1.07 | 0.61 | 0.56 | 1.88 | 1.84 | 2.05 | 2.43 | 2.05 |
| 44 | - | - | - | - | - | - | 3.96 | 3.96 |
| 45 | 1.61 | - | - | 0.44 | -0.14 | - | - | -0.15 |
| 46 | - | 1.27 | 0.89 | - | 2.91 | 2.46 | 2.51 | 2.62 |
| 47 | 1.29 | - | - | -0.05 | -0.31 | - | - | -0.18 |
| 48 | 1.23 | 1.17 | - | 0.52 | 0.32 | -0.03 | - | 0.27 |
| Σ | 23.79 | 22.14 | 26.42 | | | | | |
| | N=23 | N=27 | N=28 | | | | | |
| $\bar{X}$ | 1.03 | 0.82 | 0.94 | | | | | |

Kategoriengrenzen:

$t_1 = 0$
$t_2 = 1.03$
$t_3 = 1.85$
$t_4 = 2.79$

Abb. 2 : Graphische Darstellung der Skalenwerte der Items aufgrund des Expertenratings

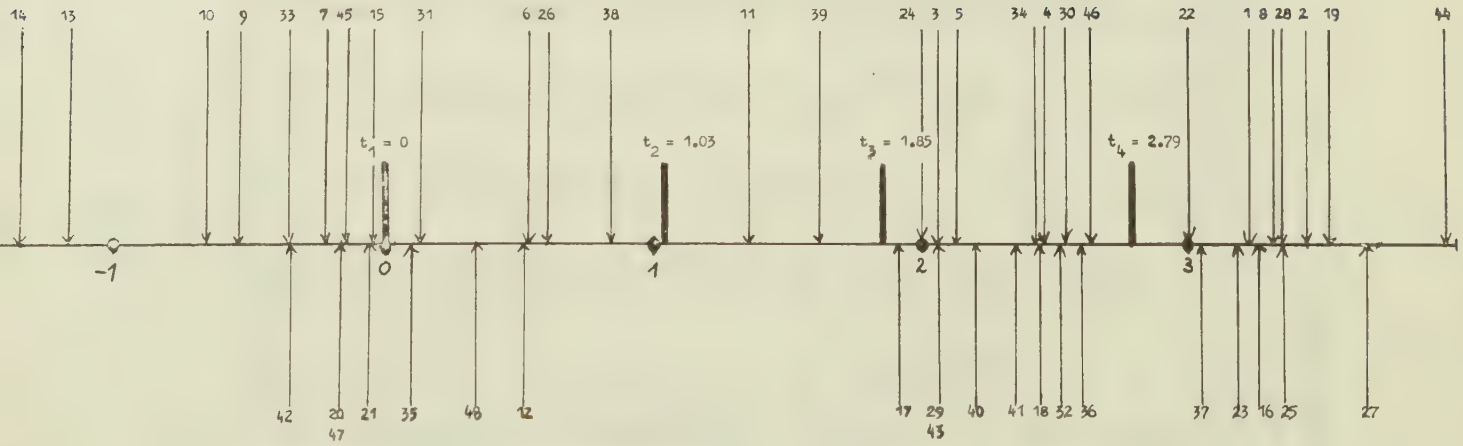


Abbildung 2 veranschaulicht das Ergebnis des Expertenratings. Ein wesentlicher Vorteil dieser Methode liegt darin, daß die Abstände zwischen den Items entlang der Skala interpretierbar sind. Nach Meinung der Experten trägt Item 44 "Bei unserem Lehrer dürfen wir offen unsere Meinung sagen" am meisten und Item 13 "Der Lehrer sagt, daß er Fragen von Schülern für dumm hält" am wenigsten zu einem guten Klassenklima bei. Diese beiden Items liegen am weitesten auseinander. Auffällig ist, daß sich in den Extremen der Skala die meisten Items scharen, während der Mittelbereich der Skala relativ unbesetzt geblieben ist, man konnte nahezu von einer zweigipfligen Verteilung sprechen, wobei ins rechte Extrem Unterstützungs- bzw. Lohnitems und ins linke Extrem Strenge- bzw. Strafitems fallen. Das Expertenrating trennt also die Items in zwei gut unterscheidbare Hälften. Dem Expertenrating zufolge begünstigen unterstützende und belohnende Verhaltensweisen Lernatmosphäre und Klassenklima in hohem Ausmaß; Tadel, Strafe, Blamage und Strenge hingegen kaum. Bei der späteren Itemselektion werden wir insbesondere Items aus den Extrempositionen berücksichtigen.

### .3.5.3. Bestimmung der Item - Validitätskoeffizienten hinsichtlich des Kriteriums soziale **E r w ü n s c h t h e i t**

Um zu kontrollieren, in welchen Items zur Lehrerbeurteilung sich die Tendenz zu sozial erwünschtem Antwortverhalten manifestiert, erfolgte eine Validierung des Schülerfragebogens am Lügenscore (vergl. Fragestellung 5). Zuvor jedoch wurde noch einmal (getrennt für die Gesamt- und Sonderschülerstichprobe) der Lügenfragebogen von ASCHERSLEBEN (1971) auf seine Reliabilität hin überprüft, weil wir aus Ökonomiegründen den Antwortmodus der Items des Lügenfragebogens verändert hatten. Wir

hatten die von ASCHERSLEBEN bezüglich der Items verwendete siebenstufige durch eine vierstufige Anwortskala ersetzt. Dennoch ist die Reliabilitätsschätzung - bezogen auf die Gesamtstichprobe - nach KUDER - RICHARDSON 21 mit 0,82 noch sehr hoch. ASCHERSLEBEN (1971, a.a.O.), der mit einem siebenstufigen Antwortmodus arbeitete, gibt als Reliabilitätsschätzung 0,84 an. Bezüglich der Sonderschülerstichprobe erhalten wir für den Lügenfragebogen immerhin noch eine Reliabilität von 0,74.

Um zu überprüfen, inwieweit die Items zur Lehrerbeurteilung sozial erwünschtes Anwortverhalten der Schüler erfassen, werden die Korrelationen zwischen den Items und dem Summenscore des Lügenfragebogens ermittelt. Korreliert werden die Itemmittelwerte (bezogen auf die jeweilige Klasse als Meßwertträger) mit dem durchschnittlichen Klassenlügenscore über die Anzahl der Klassen (N = 40). Tabelle 12 enthält die Itemmittelwerte und den durchschnittlichen Lügenscore für jede Klasse, außerdem die Item-Validierungskoeffizienten $r_{iL}$.

7.3.5.4. Varianzanalytische Überprüfung der Itembrauchbarkeit

Für die einzelnen beobachtbaren Verhaltensmerkmale (Items) hatten wir in Kapitel 1 die Forderung erhoben, daß den Schülern als Beobachter die Einstufung des beobachtbaren Verhaltens "leicht" gemacht werden müsse. Die Schüler einer bestimmten Klasse sollten sich bezüglich ihrer Urteile, d.h. bezüglich der quantitativen Ausprägung des durch ein Item erfaßten Lehrerverhaltens "einig" sein. Ein Item kann nur dann zwischen Lehrern differenzieren wenn die Streuung der Schülerurteile innerhalb der einzelnen Klassen gering und zwischen den Klassen groß ist. Varianzanalytisch können wir also eine gesamte Itemvarianz

Tab. 12
Itemmittelwerte, Lügenscores und Item-Validitätskoeffizienten $r_{it}$

## Itemmittelwerte pro Klasse

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3.50 | 3.36 | 3.00 | 3.43 | 2.76 | 3.76 | 3.67 | 3.00 | 3.29 | 3.27 | 2.96 | 3.38 | 3.19 | 3.67 | 3.12 | 3.46 | 3.24 | 3.50 | 2.88 | 2.64 |
| 3.50 | 3.79 | 3.28 | 3.43 | 3.71 | 3.70 | 3.47 | 3.33 | 3.07 | 3.53 | 3.75 | 3.85 | 3.69 | 3.61 | 4.00 | 3.23 | 3.80 | 3.80 | 3.62 | 3.00 |
| 3.00 | 3.71 | 3.72 | 3.64 | 3.76 | 3.70 | 3.60 | 4.00 | 3.50 | 3.40 | 3.71 | 3.62 | 3.38 | 3.61 | 3.92 | 3.77 | 3.12 | 3.20 | 3.29 | 2.45 |
| 2.25 | 3.07 | 2.78 | 3.50 | 3.29 | 3.20 | 2.87 | 3.07 | 3.14 | 3.00 | 3.46 | 3.31 | 3.56 | 3.17 | 3.88 | 4.00 | 2.68 | 2.90 | 2.75 | 2.68 |
| 3.33 | 3.71 | 2.94 | 3.71 | 3.35 | 3.60 | 3.73 | 3.33 | 3.64 | 3.47 | 3.58 | 3.54 | 3.56 | 3.94 | 3.88 | 4.00 | 3.80 | 3.63 | 3.50 | 2.95 |
| 4.00 | 4.00 | 3.56 | 3.86 | 3.82 | 3.80 | 3.80 | 3.93 | 3.71 | 3.27 | 3.80 | 3.62 | 3.69 | 3.72 | 3.35 | 4.00 | 3.96 | 3.77 | 3.71 | 3.59 |
| 2.33 | 2.43 | 2.78 | 1.93 | 2.24 | 2.00 | 2.00 | 2.93 | 2.14 | 2.40 | 2.33 | 2.38 | 2.63 | 2.17 | 2.08 | 2.15 | 2.24 | 2.30 | 2.75 | 2.36 |
| 3.25 | 3.79 | 3.11 | 3.72 | 3.65 | 3.10 | 3.80 | 2.87 | 3.29 | 3.42 | 3.33 | 3.46 | 3.31 | 3.89 | 3.46 | 3.31 | 3.48 | 3.77 | 3.88 | 2.64 |
| 3.42 | 3.29 | 2.83 | 3.50 | 2.94 | 3.20 | 3.20 | 2.20 | 3.29 | 2.87 | 3.08 | 3.15 | 2.88 | 3.44 | 3.66 | 3.88 | 3.08 | 2.93 | 2.83 | 2.36 |
| 1.67 | 2.14 | 2.56 | 1.86 | 2.88 | 2.50 | 1.93 | 3.27 | 1.64 | 1.87 | 2.29 | 2.23 | 1.63 | 2.50 | 2.54 | 1.85 | 1.56 | 1.50 | 2.17 | 1.55 |
| 3.08 | 3.79 | 3.28 | 3.64 | 3.82 | 3.30 | 3.13 | 3.53 | 3.79 | 3.20 | 3.71 | 3.54 | 3.44 | 3.56 | 2.73 | 2.54 | 3.40 | 3.23 | 3.50 | 2.91 |
| 3.50 | 3.50 | 3.11 | 3.50 | 3.18 | 3.60 | 3.80 | 3.40 | 3.14 | 3.33 | 3.42 | 3.46 | 3.25 | 3.44 | 2.77 | 3.48 | 3.52 | 3.17 | 3.50 | 3.14 |
| 1.17 | 1.64 | 2.00 | 1.79 | 1.82 | 2.00 | 1.33 | 1.73 | 1.64 | 1.93 | 1.88 | 1.54 | 2.00 | 1.61 | 1.23 | 1.46 | 1.88 | 2.10 | 2.42 | 1.82 |
| 2.33 | 1.14 | 2.00 | 1.43 | 2.24 | 1.70 | 1.40 | 1.87 | 1.21 | 1.73 | 1.58 | 1.69 | 2.00 | 1.06 | 1.54 | 1.77 | 1.20 | 1.17 | 2.58 | 1.68 |
| 2.25 | 1.43 | 2.00 | 1.71 | 1.41 | 1.20 | 1.87 | 1.80 | 1.36 | 1.80 | 1.50 | 1.38 | 1.31 | 2.17 | 1.38 | 1.69 | 1.64 | 1.80 | 2.29 | 2.05 |
| 1.92 | 1.07 | 2.17 | 1.14 | 1.41 | 1.40 | 1.53 | 1.80 | 1.50 | 1.67 | 1.71 | 1.46 | 1.75 | 1.22 | 1.04 | 1.23 | 1.24 | 1.07 | 1.58 | 1.41 |
| 2.75 | 3.14 | 3.06 | 3.14 | 3.18 | 2.40 | 3.33 | 2.20 | 3.00 | 2.93 | 3.04 | 2.92 | 2.81 | 3.28 | 3.15 | 2.77 | 3.64 | 3.57 | 3.13 | 2.91 |
| 3.75 | 3.64 | 3.28 | 3.64 | 2.94 | 3.20 | 4.00 | 3.27 | 3.36 | 3.40 | 3.75 | 3.85 | 3.13 | 3.67 | 3.27 | 2.46 | 3.20 | 3.23 | 3.63 | 2.73 |
| 3.50 | 3.14 | 2.33 | 2.93 | 2.29 | 3.20 | 3.40 | 2.87 | 2.50 | 3.00 | 3.21 | 3.15 | 2.75 | 3.11 | 3.19 | 3.40 | 2.40 | 2.47 | 2.54 | 2.82 |
| 2.42 | 2.07 | 3.17 | 2.93 | 1.82 | 2.22 | 2.53 | 3.20 | 2.00 | 2.53 | 2.92 | 2.31 | 2.44 | 1.94 | 2.27 | 3.31 | 2.28 | 2.33 | 2.38 | 2.32 |
| 3.17 | 3.29 | 3.06 | 3.71 | 3.18 | 2.80 | 3.67 | 3.60 | 3.36 | 2.73 | 3.25 | 3.15 | 2.88 | 3.61 | 3.23 | 3.15 | 3.48 | 3.63 | 2.75 | 2.82 |
| 3.42 | 3.71 | 3.22 | 3.57 | 3.53 | 2.92 | 3.00 | 2.60 | 3.00 | 3.53 | 3.33 | 3.08 | 3.06 | 3.72 | 3.23 | 3.31 | 3.20 | 2.87 | 3.06 | 2.64 |
| 3.17 | 3.50 | 3.39 | 3.29 | 3.12 | 3.50 | 3.53 | 3.27 | 2.21 | 3.80 | 3.71 | 3.85 | 3.31 | 3.50 | 2.85 | 3.92 | 3.48 | 3.37 | 3.50 | 2.82 |
| 3.33 | 3.57 | 3.44 | 3.64 | 2.59 | 3.50 | 3.60 | 3.47 | 3.14 | 3.40 | 3.50 | 3.46 | 3.13 | 3.44 | 2.15 | 3.77 | 3.12 | 2.47 | 2.58 | 2.55 |
| 3.17 | 3.14 | 3.06 | 3.36 | 2.82 | 3.80 | 3.40 | 2.73 | 3.00 | 3.40 | 3.58 | 3.46 | 2.94 | 3.44 | 3.00 | 3.62 | 3.04 | 3.33 | 3.08 | 2.41 |
| 2.83 | 2.57 | 2.89 | 2.86 | 2.18 | 2.80 | 2.40 | 3.40 | 1.71 | 3.00 | 2.67 | 2.54 | 2.81 | 2.22 | 2.46 | 3.00 | 2.36 | 2.00 | 2.75 | 2.50 |
| 3.50 | 3.14 | 2.67 | 3.64 | 3.00 | 2.70 | 3.00 | 2.73 | 3.07 | 3.13 | 3.42 | 2.77 | 2.81 | 3.33 | 3.69 | 3.00 | 3.40 | 3.40 | 2.92 | 3.09 |
| 3.58 | 3.71 | 3.00 | 3.43 | 2.88 | 3.30 | 3.87 | 3.00 | 3.14 | 3.53 | 3.67 | 3.15 | 3.19 | 3.78 | 3.65 | 3.46 | 3.96 | 3.53 | 3.25 | 3.59 |
| 2.83 | 3.07 | 2.56 | 3.50 | 2.59 | 2.30 | 3.07 | 3.60 | 3.00 | 3.13 | 2.75 | 2.54 | 3.06 | 2.67 | 3.12 | 2.62 | 3.60 | 3.73 | 3.33 | 2.86 |
| 2.00 | 2.07 | 2.67 | 2.21 | 2.59 | 1.80 | 1.80 | 2.87 | 2.00 | 2.00 | 2.04 | 1.92 | 2.44 | 1.83 | 1.65 | 2.62 | 2.00 | 1.80 | 2.17 | 2.23 |
| 2.58 | 2.86 | 3.56 | 2.71 | 2.47 | 3.10 | 2.67 | 3.07 | 2.43 | 3.07 | 2.58 | 1.92 | 2.88 | 2.28 | 2.42 | 2.77 | 2.40 | 2.43 | 3.04 | 3.00 |
| 3.42 | 3.21 | 3.44 | 3.36 | 3.82 | 3.70 | 3.87 | 3.27 | 3.50 | 3.33 | 3.42 | 3.46 | 3.44 | 3.94 | 3.58 | 3.46 | 3.64 | 3.67 | 3.67 | 3.55 |
| 1.83 | 2.29 | 3.28 | 2.64 | 1.88 | 1.90 | 1.47 | 2.33 | 1.50 | 1.80 | 1.83 | 1.46 | 1.75 | 1.33 | 1.35 | 2.69 | 1.96 | 1.67 | 2.08 | 1.77 |
| 1.67 | 1.07 | 2.06 | 2.14 | 2.12 | 1.50 | 1.40 | 2.20 | 1.64 | 1.60 | 2.00 | 2.15 | 1.81 | 1.33 | 3.15 | 1.38 | 1.60 | 1.63 | 2.04 | 2.14 |
| 2.50 | 2.00 | 2.61 | 2.29 | 2.24 | 2.10 | 2.00 | 2.73 | 2.50 | 1.93 | 2.50 | 2.08 | 2.56 | 1.94 | 2.31 | 2.69 | 1.72 | 1.63 | 3.00 | 2.95 |
| 2.33 | 1.86 | 2.33 | 1.86 | 1.76 | 1.90 | 2.33 | 2.33 | 2.14 | 1.67 | 2.33 | 1.62 | 2.19 | 1.61 | 1.77 | 2.23 | 1.84 | 1.77 | 2.33 | 2.41 |
| 3.75 | 3.79 | 3.72 | 3.71 | 3.41 | 3.80 | 3.80 | 3.67 | 3.64 | 3.60 | 3.92 | 3.62 | 3.56 | 3.56 | 3.88 | 3.92 | 3.84 | 3.60 | 3.67 | 3.45 |
| 1.67 | 2.57 | 2.67 | 2.07 | 2.82 | 2.10 | 2.87 | 2.60 | 2.57 | 2.40 | 2.50 | 2.69 | 2.19 | 2.11 | 2.58 | 2.08 | 1.96 | 2.03 | 2.50 | 2.05 |
| 1.92 | 1.50 | 2.33 | 1.79 | 2.00 | 2.40 | 1.67 | 2.20 | 1.93 | 1.60 | 1.92 | 1.85 | 1.88 | 1.50 | 1.92 | 2.62 | 1.64 | 1.50 | 1.96 | 1.86 |
| 3.58 | 3.79 | 3.28 | 3.43 | 3.65 | 3.90 | 3.67 | 2.67 | 3.93 | 3.53 | 3.46 | 3.38 | 3.69 | 3.67 | 3.96 | 3.23 | 3.96 | 3.60 | 3.96 | 3.64 |
| 1.67 | 1.21 | 1.94 | 1.57 | 1.29 | 2.10 | 1.60 | 1.87 | 1.64 | 1.60 | 1.46 | 1.62 | 1.75 | 1.39 | 1.27 | 1.69 | 1.16 | 1.30 | 1.71 | 1.64 |
| 3.33 | 3.50 | 3.33 | 2.21 | 1.82 | 2.20 | 1.73 | 2.07 | 1.86 | 1.53 | 2.27 | 1.77 | 3.31 | 1.67 | 1.08 | 2.54 | 1.84 | 1.57 | 2.13 | 2.50 |
| 2.00 | 1.57 | 2.00 | 1.57 | 1.71 | 1.70 | 1.93 | 2.27 | 1.64 | 1.53 | 1.83 | 1.31 | 1.56 | 1.33 | 1.12 | 2.15 | 1.24 | 1.10 | 1.71 | 1.59 |
| 2.00 | 1.50 | 2.39 | 1.50 | 1.53 | 1.70 | 1.20 | 1.93 | 1.64 | 1.47 | 1.79 | 1.54 | 1.50 | 1.22 | 1.12 | 1.77 | 1.16 | 1.17 | 1.58 | 1.41 |
| 2.33 | 1.71 | 2.28 | 1.93 | 1.94 | 1.90 | 1.93 | 2.20 | 1.50 | 2.13 | 1.71 | 1.69 | 2.06 | 1.67 | 2.15 | 1.62 | 1.88 | 1.90 | 2.46 | 2.32 |
| 2.83 | 1.43 | 2.39 | 1.71 | 2.12 | 2.60 | 1.80 | 2.80 | 1.86 | 2.13 | 1.75 | 2.23 | 2.06 | 1.94 | 1.38 | 2.38 | 1.72 | 1.73 | 2.38 | 2.00 |
| 2.92 | 2.21 | 3.17 | 2.21 | 2.24 | 2.80 | 2.07 | 3.47 | 2.21 | 2.74 | 2.33 | 2.08 | 2.75 | 2.28 | 2.04 | 2.31 | 1.92 | 1.87 | 2.75 | 2.68 |
| 2.17 | 1.93 | 3.06 | 1.79 | 1.94 | 1.70 | 2.27 | 2.67 | 1.93 | 1.93 | 2.13 | 2.40 | 2.31 | 1.78 | 2.38 | 2.08 | 1.76 | 1.60 | 2.79 | 2.09 |
| 80,8 | 85,7 | 80,9 | 82,1 | 86,8 | 78,5 | 83,3 | 79,8 | 84,9 | 81,9 | 82,3 | 81,7 | 81,1 | 79,3 | 84,2 | 80,0 | 82,5 | 82,6 | 81,5 | 77,3 |

## Fortsetzung von Tab. 11

| Item Nr. | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2,69 | 2,97 | 3,00 | 2,92 | 3,44 | 2,76 | 3,65 | 2,50 | 3,45 | 2,89 | 1,96 | 2,78 | 3,61 | 2,71 | 2,14 | 2,43 | 3,15 | 3,27 | 3,11 | 3,07 | |
| 2 | 3,39 | 3,58 | 3,56 | 3,13 | 3,92 | 2,90 | 4,00 | 2,88 | 3,82 | 3,78 | 3,84 | 3,95 | 3,48 | 3,40 | 3,03 | 2,67 | 3,93 | 3,57 | 3,54 | 3,84 | |
| 3 | 3,14 | 3,32 | 3,82 | 3,10 | 3,40 | 2,66 | 3,35 | 3,05 | 3,41 | 3,19 | 3,30 | 3,30 | 3,32 | 3,05 | 3,46 | 3,00 | 3,67 | 2,73 | 3,71 | 3,21 | |
| 4 | 1,89 | 2,97 | 3,06 | 1,78 | 3,48 | 1,84 | 3,91 | 2,55 | 3,36 | 2,51 | 1,30 | 2,49 | 3,68 | 2,14 | 3,91 | 3,05 | 3,63 | 2,49 | 3,79 | 3,72 | |
| 5 | 3,62 | 3,75 | 3,67 | 3,58 | 3,84 | 3,33 | 4,00 | 2,32 | 3,86 | 3,54 | 3,76 | 3,51 | 3,42 | 3,29 | 2,42 | 2,62 | 3,67 | 3,54 | 3,21 | 3,66 | |
| 6 | 3,92 | 3,64 | 4,00 | 3,95 | 3,92 | 4,00 | 3,87 | 3,73 | 2,55 | 3,97 | 3,96 | 3,62 | 3,06 | 3,52 | 3,83 | 3,57 | 3,64 | 2,74 | 3,83 | 3,69 | |
| 7 | 2,32 | 1,92 | 2,31 | 2,55 | 2,16 | 2,81 | 1,91 | 3,24 | 2,00 | 2,19 | 3,00 | 2,32 | 2,06 | 2,62 | 3,09 | 2,29 | 2,30 | 1,73 | 2,35 | 2,69 | |
| 8 | 3,47 | 3,75 | 3,58 | 3,83 | 3,60 | 2,62 | 3,87 | 2,05 | 3,50 | 3,49 | 3,32 | 3,46 | 2,97 | 3,33 | 2,74 | 2,71 | 3,22 | 3,17 | 2,99 | 3,66 | |
| 9 | 2,86 | 3,00 | 2,94 | 2,72 | 2,96 | 2,24 | 3,61 | 2,32 | 3,32 | 3,24 | 3,04 | 3,00 | 2,77 | 2,71 | 2,46 | 1,67 | 3,37 | 3,03 | 2,11 | 3,12 | |
| 10 | 2,30 | 2,36 | 2,17 | 2,15 | 1,72 | 2,57 | 1,65 | 2,86 | 2,05 | 2,35 | 2,72 | 2,00 | 1,55 | 2,24 | 2,49 | 1,52 | 2,30 | 1,69 | 3,02 | 1,90 | |
| 11 | 3,03 | 3,08 | 3,53 | 3,05 | 3,32 | 2,81 | 3,35 | 2,30 | 3,23 | 3,22 | 2,76 | 3,30 | 3,13 | 2,86 | 2,63 | 3,22 | 3,15 | 2,97 | 2,84 | 3,17 | |
| 12 | 3,22 | 3,44 | 3,50 | 3,28 | 3,56 | 3,19 | 3,83 | 2,68 | 3,14 | 3,38 | 2,56 | 3,19 | 2,87 | 2,67 | 2,80 | 2,67 | 2,74 | 2,83 | 2,84 | 3,18 | |
| 13 | 1,67 | 1,39 | 1,78 | 1,58 | 1,76 | 2,90 | 1,39 | 1,64 | 1,91 | 1,80 | 1,20 | 1,92 | 1,39 | 2,19 | 2,94 | 2,05 | 1,33 | 1,47 | 1,80 | 1,59 | |
| 14 | 1,61 | 1,06 | 1,92 | 2,08 | 1,24 | 2,67 | 1,22 | 3,00 | 1,64 | 1,68 | 1,56 | 1,78 | 1,45 | 1,71 | 3,37 | 1,81 | 1,41 | 1,17 | 2,54 | 1,68 | |
| 15 | 2,08 | 1,39 | 1,83 | 2,15 | 1,44 | 3,05 | 1,52 | 3,32 | 1,45 | 1,46 | 2,16 | 2,54 | 2,26 | 1,67 | 3,57 | 2,43 | 1,26 | 1,27 | 2,29 | 1,95 | |
| 16 | 1,58 | 1,19 | 1,58 | 1,58 | 1,16 | 2,33 | 1,04 | 1,82 | 1,14 | 1,19 | 1,80 | 1,35 | 1,13 | 1,24 | 2,34 | 1,62 | 1,13 | 1,17 | 1,34 | 1,38 | |
| 17 | 3,28 | 3,39 | 3,00 | 2,95 | 3,56 | 2,81 | 3,65 | 2,18 | 3,05 | 3,22 | 2,20 | 3,03 | 3,13 | 2,95 | 1,89 | 2,47 | 2,04 | 3,50 | 2,71 | 3,14 | |
| 18 | 3,33 | 3,36 | 3,33 | 3,47 | 3,60 | 2,80 | 3,83 | 3,59 | 3,36 | 3,65 | 3,44 | 3,51 | 3,94 | 3,48 | 3,43 | 2,74 | 3,22 | 3,43 | 3,18 | 3,42 | |
| 19 | 2,86 | 2,92 | 2,56 | 2,22 | 3,16 | 2,24 | 3,74 | 2,23 | 2,95 | 2,35 | 1,36 | 1,89 | 2,87 | 2,10 | 1,48 | 1,12 | 1,85 | 3,73 | 1,36 | 2,79 | |
| 20 | 2,67 | 1,78 | 2,53 | 2,15 | 2,56 | 3,33 | 2,35 | 3,27 | 1,82 | 2,39 | 2,72 | 1,76 | 2,32 | 2,57 | 3,26 | 2,41 | 2,25 | 2,23 | 2,34 | 2,41 | |
| 21 | 3,52 | 3,69 | 3,50 | 3,58 | 3,52 | 2,48 | 3,39 | 1,86 | 2,95 | 3,49 | 3,24 | 3,19 | 3,06 | 3,38 | 2,51 | 2,77 | 3,47 | 3,75 | 3,07 | 3,45 | |
| 22 | 3,17 | 3,50 | 3,03 | 2,95 | 3,12 | 2,38 | 2,96 | 2,41 | 2,95 | 3,27 | 2,84 | 3,24 | 2,74 | 3,00 | 2,71 | 2,86 | 2,67 | 2,97 | 3,14 | 2,98 | |
| 23 | 3,25 | 3,50 | 3,72 | 3,13 | 3,60 | 2,95 | 3,61 | 3,09 | 3,27 | 3,51 | 2,96 | 3,24 | 3,26 | 3,10 | 2,85 | 2,95 | 3,21 | 3,57 | 3,04 | 3,41 | |
| 24 | 3,44 | 3,20 | 3,75 | 3,47 | 3,60 | 2,67 | 3,09 | 1,45 | 2,95 | 3,03 | 2,48 | 2,70 | 2,45 | 2,62 | 2,23 | 2,40 | 3,37 | 2,90 | 2,33 | 3,21 | |
| 25 | 3,06 | 3,61 | 3,30 | 3,25 | 3,48 | 2,81 | 3,70 | 1,86 | 3,00 | 2,97 | 3,06 | 3,22 | 2,71 | 2,71 | 2,31 | 1,80 | 2,78 | 2,84 | 2,67 | 3,34 | |
| 26 | 2,47 | 1,97 | 2,53 | 2,75 | 2,60 | 3,24 | 2,04 | 3,36 | 1,91 | 2,43 | 2,80 | 1,86 | 2,13 | 2,81 | 3,60 | 2,05 | 2,44 | 2,11 | 2,41 | 2,59 | |
| 27 | 3,25 | 3,61 | 3,56 | 3,35 | 3,48 | 2,86 | 3,96 | 2,14 | 3,57 | 3,17 | 3,48 | 3,49 | 3,74 | 2,86 | 1,74 | 2,22 | 2,52 | 3,62 | | | |
| 28 | 3,31 | 3,89 | 3,42 | 3,42 | 3,84 | 2,48 | 3,91 | 1,73 | 3,82 | 3,57 | 3,28 | 3,84 | 3,97 | 3,10 | 2,29 | 2,00 | 3,50 | 3,00 | 2,68 | 3,76 | |
| 29 | 3,14 | 3,52 | 3,31 | 3,33 | 3,44 | 2,52 | 3,87 | 1,82 | 2,50 | 3,27 | 3,48 | 3,32 | 3,32 | 2,66 | 1,91 | 2,48 | 2,90 | 3,57 | 3,04 | 3,83 | |
| 30 | 1,78 | 1,50 | 2,06 | 2,08 | 1,60 | 2,71 | 1,57 | 2,86 | 1,45 | 1,73 | 2,08 | 1,84 | 1,87 | 1,86 | 3,09 | 2,02 | 1,88 | 1,57 | 2,57 | 2,07 | |
| 31 | 3,00 | 2,11 | 3,00 | 3,10 | 2,92 | 3,29 | 2,22 | 3,27 | 2,14 | 2,68 | 3,36 | 2,92 | 2,74 | 3,00 | 3,69 | 3,19 | 2,45 | 2,13 | 2,54 | 2,72 | |
| 32 | 3,33 | 3,72 | 3,31 | 3,33 | 3,56 | 2,67 | 3,78 | 2,32 | 3,64 | 3,49 | 3,52 | 3,32 | 3,74 | 3,57 | 2,77 | 2,19 | 3,24 | 3,67 | 3,00 | 3,74 | |
| 33 | 2,44 | 2,19 | 2,25 | 1,97 | 2,80 | 2,76 | 1,83 | 3,41 | 1,45 | 1,86 | 1,64 | 3,00 | 1,94 | 1,57 | 2,07 | 1,38 | 1,82 | 1,59 | 2,74 | 2,03 | |
| 34 | 1,47 | 1,64 | 2,08 | 1,53 | 1,16 | 2,40 | 1,22 | 3,45 | 1,08 | 1,81 | 3,72 | 1,59 | 1,74 | 2,10 | 2,77 | 1,71 | 2,07 | 1,50 | 1,86 | 2,10 | |
| 35 | 3,17 | 1,72 | 3,06 | 3,03 | 1,92 | 3,81 | 1,87 | 3,50 | 1,59 | 2,19 | 2,12 | 1,68 | 1,45 | 2,57 | 2,91 | 2,33 | 2,63 | 1,87 | 2,67 | 2,21 | |
| 36 | 2,39 | 1,86 | 2,11 | 2,38 | 1,56 | 3,05 | 1,48 | 3,45 | 1,59 | 1,89 | 2,48 | 1,57 | 1,48 | 2,76 | 3,14 | 2,48 | 1,67 | 1,63 | 2,25 | 1,97 | |
| 37 | 3,67 | 3,86 | 3,81 | 3,70 | 3,80 | 2,86 | 4,00 | 3,00 | 3,36 | 3,81 | 3,64 | 3,81 | 3,71 | 3,57 | 3,34 | 2,45 | 3,63 | 3,50 | 3,64 | 3,83 | |
| 38 | 2,69 | 2,67 | 2,97 | 2,78 | 2,76 | 2,33 | 2,26 | 2,60 | 2,14 | 2,65 | 3,32 | 2,30 | 1,90 | 2,76 | 2,60 | 1,33 | 3,04 | 2,21 | 2,89 | 2,65 | |
| 39 | 1,78 | 1,53 | 1,94 | 1,95 | 1,92 | 2,29 | 1,43 | 2,68 | 1,66 | 1,76 | 2,16 | 1,76 | 1,71 | 1,52 | 2,22 | 1,44 | 1,47 | 1,86 | 1,91 | 1,81 | |
| 40 | 3,50 | 4,00 | 3,89 | 3,72 | 3,92 | 3,14 | 4,00 | 3,36 | 3,86 | 3,89 | 3,72 | 3,97 | 3,87 | 3,81 | 3,14 | 3,00 | 4,00 | 3,80 | 3,86 | 4,00 | |
| 41 | 1,86 | 1,08 | 1,56 | 1,69 | 1,20 | 2,33 | 1,00 | 2,00 | 1,14 | 1,41 | 1,52 | 1,24 | 1,39 | 1,48 | 2,43 | 1,15 | 1,50 | 1,44 | | | |
| 42 | 3,11 | 2,00 | 2,81 | 2,70 | 2,64 | 3,48 | 2,39 | 3,68 | 2,23 | 2,46 | 1,72 | 2,95 | 1,94 | 1,76 | 3,91 | 3,19 | 2,63 | 2,30 | 3,11 | 3,14 | |
| 43 | 1,72 | 1,36 | 1,56 | 1,58 | 1,32 | 2,29 | 1,22 | 2,00 | 1,27 | 1,41 | 1,56 | 1,57 | 1,10 | 1,24 | 2,31 | 1,50 | 1,15 | 1,27 | 1,50 | 1,55 | |
| 44 | 1,56 | 1,11 | 1,53 | 1,42 | 1,44 | 2,24 | 1,17 | 2,64 | 1,09 | 1,27 | 1,44 | 1,16 | 1,13 | 1,90 | 3,06 | 2,29 | 1,41 | 1,13 | 2,14 | 1,21 | |
| 45 | 2,17 | 1,78 | 2,56 | 2,55 | 2,32 | 2,71 | 1,83 | 2,48 | 2,18 | 2,00 | 2,32 | 2,46 | 2,23 | 2,43 | 2,71 | 2,81 | 2,11 | 2,03 | 2,39 | 2,24 | |
| 46 | 2,17 | 1,53 | 2,17 | 2,38 | 1,64 | 2,62 | 1,78 | 3,09 | 2,05 | 1,92 | 2,32 | 1,89 | 1,42 | 2,38 | 3,23 | 2,52 | 2,19 | 1,51 | 2,04 | 1,91 | |
| 47 | 2,83 | 2,06 | 2,64 | 2,50 | 2,08 | 3,62 | 1,96 | 3,59 | 2,05 | 2,22 | 2,80 | 2,03 | 1,84 | 2,90 | 3,74 | 3,10 | 2,15 | 1,67 | 2,29 | 2,11 | |
| 48 | 2,31 | 1,75 | 2,36 | 2,47 | 2,44 | 2,76 | 1,85 | 3,36 | 2,14 | 2,08 | 2,84 | 2,24 | 1,87 | 2,29 | 3,09 | 2,82 | 2,46 | 1,77 | 2,61 | 2,62 | |
| Z-Score | 75,5 | 83,8 | 86,7 | 82,2 | 83,5 | 85,2 | 78,4 | 80,0 | 81,4 | 77,4 | 83,8 | 78,3 | 82,0 | 76,3 | 76,2 | 80,7 | 73,5 | 77,1 | 78,8 | 83,9 | |

(beruhend auf 937 Schülerurteilen aus 40 Klassen als Meßwerte) aufspalten in:

Varianz zwischen den Klassen (Lehrern) + Varianz innerhalb der Klassen + Fehlervarianz.

Die Frage, ob sich die 40 Klassenmittelwerte bezüglich eines Items signifikant voneinander unterscheiden, läßt sich mit Hilfe einer einfaktoriellen Varianzanalyse beantworten, in der die 40 Klassen die Stufen der unabhängigen Variablen und ein jeweiliges Item die abhängige Variable darstellen. Hierbei handelt es sich um ein random model. Bei 48 Items (48 abhängige Variablen) hätten wir 48 Varianzanalysen zu rechnen.

Sinnvoll erschien uns, die Items daraufhin zu überprüfen, ob sie zwischen den Klassen innerhalb der drei Schultypen gleichermaßen gut trennen. Demzufolge hätten wir für die Haupt-, Real- und Sonderschülerstichprobe je 48 Varianzanalysen zu rechnen. Für die Haupt- und Sonderschülerstichprobe sind pro Item (Variable) die Unterschiede zwischen je 15 Mittelwerten (entsprechend 15 Klassen), bei der Realschullehrerstichprobe die Unterschiede zwischen 10 Mittelwerten (entsprechend 10 Klassen) auf Signifikanz zu prüfen. Als Ergebnisse dieser Varianzanalysen weist Tabelle 13 a unter anderem die F - Prüfgrößen einschließlich deren Signifikanzniveaus aus. Erwartungsgemäß fielen die F - Werte für die Sonderschülerstichprobe niedriger aus. Während im Hinblick auf die Haupt- und Realschülerstichprobe alle Items hoch signifikant zwischen den Klassen trennen, sind für die Sonderschülerstichprobe einige nicht signifikante F - Tests zu verzeichnen.

Tab. 13a

Überblick über alle zur Itemselektion errechneten Kennwerte

| Item Nr. | Subtest | Gesamt-Stichprobe $r_{it}$ | Gesamt-Stichprobe p | Sonderschüler-stichprobe $r_{it}$ | Sonderschüler-stichprobe p | $r_{iL}$ | Faktorenladungen der Items für die Gesamt-stichprobe S | Gesamt-stichprobe U | Sonder-schüler-stichprobe S | Sonder-schüler-stichprobe U | F-Prüfgrößen und deren Signifikanzniveaus als Ergebnisse der Varianzanalysen zur Überprüfung der Itemmittelwertsunterschiede zwischen den Klassen der Sonder-, | p' | Haupt- und | p' | Realschulen | p' | Skalenwerte der Items aufgrund des Expertenratings |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| * 1 | U | .49 | .67 | .43 | .76 | .12 | .34 | .37 | .27 | .31 | 2.53 | 0.002 | 7.13 | 0.0000 | 18.97 | 0.0000 | 3.23 |
| 2 | U | .48 | .85 | .31 | .84 | .12 | .33 | .38 | .26 | .18 | 1.51 | 0.107 | 13.35 | 0.0000 | 12.22 | 0.0000 | 3.43 |
| 3 | U | .25 | .78 | .15 | .87 | .32 | -.09 | .29 | .00 | .05 | 1.60 | 0.080 | 6.78 | 0.0000 | 4.89 | 0.0000 | 2.05 |
| 4 | U | .27 | .65 | .36 | .73 | .05 | .12 | .17 | .25 | .20 | 3.34 | 0.000 | 30.77 | 0.0000 | 50.77 | 0.0000 | 2.46 |
| 5 | U | .57 | .76 | .52 | .85 | .16 | .26 | .53 | .26 | .44 | 1.83 | 0.035 | 11.87 | 0.0000 | 10.58 | 0.0000 | 2.13 |
| 6 | S | .22 | .90 | .08 | .92 | .40 | -.36 | .25 | -.08 | .14 | 2.10 | 0.012 | 13.08 | 0.0000 | 16.13 | 0.0000 | 0.53 |
| * 7 | S | .56 | .45 | .46 | .44 | .18 | -.50 | -.27 | -.46 | -.20 | 1.84 | 0.033 | 7.78 | 0.0000 | 12.89 | 0.0000 | -0.21 |
| 8 | U | .59 | .74 | .43 | .81 | .44 | .22 | .56 | .19 | .40 | 1.94 | 0.023 | 12.86 | 0.0000 | 3.00 | 0.0023 | 3.31 |
| * 9 | U | .56 | .63 | .49 | .67 | .16 | .35 | .43 | .46 | .27 | 3.55 | 0.000 | 4.64 | 0.0000 | 10.93 | 0.0000 | -0.55 |
| 10 | S | .38 | .37 | .36 | .37 | .00 | .38 | .02 | .29 | .08 | 2.71 | 0.001 | 6.19 | 0.0000 | 10.90 | 0.0000 | -0.66 |
| 11 | U | .38 | .72 | .30 | .81 | .45 | .01 | .39 | -.06 | .32 | 2.52 | 0.002 | 3.45 | 0.0001 | 2.18 | 0.0228 | 1.37 |
| 12 | U | .35 | .73 | .35 | .79 | .40 | -.10 | .42 | -.13 | .45 | 0.79 | 0.679 | 2.76 | 0.0008 | 2.23 | 0.0199 | 0.52 |
| * 13 | S | .38 | .26 | .29 | .24 | .04 | -.34 | -.14 | -.29 | -.07 | 1.08 | 0.369 | 6.69 | 0.0000 | 11.68 | 0.0000 | -1.17 |
| * 14 | S | .62 | .24 | .48 | .20 | -.29 | -.55 | -.28 | -.51 | -.06 | 2.20 | 0.008 | 12.82 | 0.0000 | 24.67 | 0.0000 | -1.34 |
| * 15 | S | .59 | .31 | .44 | .22 | -.13 | .48 | .36 | .44 | .11 | 1.87 | 0.030 | 10.62 | 0.0000 | 22.33 | 0.0000 | -0.03 |
| * 16 | S | .56 | .15 | .43 | .18 | -.09 | .49 | .32 | .37 | .26 | 2.27 | 0.006 | 8.18 | 0.0000 | 11.39 | 0.0000 | 3.25 |
| 17 | U | .49 | .67 | .29 | .65 | .32 | .29 | .41 | .12 | .23 | 1.26 | 0.230 | 3.94 | 0.0000 | 9.40 | 0.0000 | 1.86 |
| 18 | U | .37 | .79 | .45 | .80 | .21 | .08 | .34 | .23 | .37 | 3.98 | 0.000 | 7.38 | 0.0000 | 4.49 | 0.0001 | 2.44 |
| * 19 | U | .52 | .52 | .42 | .65 | .12 | .18 | .44 | .15 | .35 | 2.42 | 0.003 | 4.62 | 0.0000 | 27.47 | 0.0000 | 3.52 |
| * 20 | S | .50 | .48 | .42 | .51 | -.27 | -.52 | -.07 | -.51 | .15 | 3.39 | 0.000 | 5.95 | 0.0000 | 8.20 | 0.0000 | -0.17 |
| * 21 | U | .47 | .74 | .40 | .75 | .27 | .27 | .37 | .36 | .15 | 1.10 | 0.357 | 11.29 | 0.0000 | 6.47 | 0.0000 | -0.04 |
| 22 | U | .41 | .68 | .42 | .76 | .61 | .22 | .34 | .34 | .30 | 1.76 | 0.040 | 3.37 | 0.0001 | 2.71 | 0.0050 | 3.00 |
| * 23 | U | .51 | .77 | .35 | .81 | .18 | .03 | .57 | -.10 | .50 | 4.50 | 0.000 | 4.33 | 0.0000 | 3.56 | 0.0005 | 3.19 |
| 24 | U | .52 | .65 | .51 | .79 | .25 | -.08 | .62 | -.02 | .55 | 1.93 | 0.024 | 14.72 | 0.0000 | 3.69 | 0.0004 | 2.00 |
| * 25 | U | .61 | .67 | .45 | .75 | .21 | .12 | .64 | .12 | .48 | 1.93 | 0.024 | 10.72 | 0.0000 | 7.66 | 0.0000 | 3.35 |
| 26 | S | .58 | .51 | .45 | .55 | -.23 | -.62 | -.07 | -.59 | .25 | 2.71 | 0.001 | 6.45 | 0.0000 | 14.32 | 0.0000 | 0.60 |
| * 27 | U | .60 | .73 | .54 | .69 | .00 | .35 | .51 | .24 | .48 | 1.32 | 0.193 | 7.39 | 0.0000 | 18.22 | 0.0000 | 3.66 |
| * 28 | U | .68 | .79 | .57 | .79 | .01 | .47 | .55 | .31 | .48 | 1.96 | 0.021 | 16.40 | 0.0000 | 19.68 | 0.0000 | 3.34 |
| 29 | U | .52 | .67 | .39 | .61 | .15 | .31 | .45 | .04 | .39 | 1.26 | 0.227 | 9.42 | 0.0000 | 13.05 | 0.0000 | 2.05 |
| * 30 | S | .57 | .35 | .49 | .40 | -.04 | .43 | .43 | .33 | .39 | 2.33 | 0.005 | 6.58 | 0.0000 | 11.26 | 0.0000 | 2.54 |
| * 31 | S | .60 | .59 | .51 | .57 | -.23 | -.61 | -.14 | -.57 | .03 | 3.63 | 0.000 | 9.42 | 0.0000 | 15.74 | 0.0000 | 0.13 |
| 32 | U | .58 | .79 | .48 | .83 | .14 | .42 | .45 | .47 | .29 | 1.74 | 0.049 | 8.50 | 0.0000 | 10.79 | 0.0000 | 2.51 |
| * 33 | S | .50 | .37 | .49 | .33 | .05 | -.56 | .03 | -.53 | .05 | 5.14 | 0.000 | 12.17 | 0.0000 | 12.30 | 0.0000 | -0.36 |
| 34 | S | .48 | .30 | .45 | .25 | -.28 | .36 | .30 | .36 | .27 | 2.82 | 0.000 | 23.91 | 0.0000 | 19.86 | 0.0000 | 2.43 |
| * 35 | S | .60 | .46 | .54 | .44 | -.15 | -.61 | -.14 | -.56 | -.08 | 1.22 | 0.256 | 22.67 | 0.0000 | 10.21 | 0.0000 | 0.10 |
| * 36 | S | .61 | .36 | .47 | .34 | -.17 | .52 | .33 | .38 | .21 | 1.38 | 0.163 | 11.94 | 0.0000 | 17.81 | 0.0000 | 2.60 |
| * 37 | U | .54 | .88 | .41 | .89 | .12 | .19 | .54 | .15 | .43 | 0.77 | 0.693 | 7.73 | 0.0000 | 9.57 | 0.0000 | 3.04 |
| 38 | S | .25 | .49 | .21 | .47 | .21 | .22 | -.02 | .01 | .28 | 1.43 | 0.137 | 2.92 | 0.0000 | 10.45 | 0.0000 | 0.84 |
| 39 | S | .51 | .28 | .53 | .31 | -.18 | -.39 | -.34 | -.40 | -.44 | 1.64 | 0.069 | 3.62 | 0.0000 | 3.12 | 0.0016 | 1.61 |
| 40 | U | .40 | .90 | .29 | .83 | .21 | .21 | .36 | .17 | .23 | 2.25 | 0.007 | 5.33 | 0.0000 | 7.70 | 0.0000 | 2.20 |
| 41 | S | .59 | .18 | .57 | .20 | -.18 | .47 | .43 | .44 | .44 | 1.05 | 0.397 | 7.76 | 0.0000 | 7.96 | 0.0000 | 2.35 |
| * 42 | S | .59 | .50 | .53 | .46 | -.03 | -.62 | -.09 | -.59 | .06 | 9.18 | 0.000 | 24.08 | 0.0000 | 33.14 | 0.0000 | -0.33 |
| 43 | S | .60 | .18 | .56 | .24 | .02 | .50 | .35 | .43 | .43 | 1.55 | 0.092 | 5.12 | 0.0000 | 8.12 | 0.0000 | 2.05 |
| * 44 | S | .68 | .18 | .57 | .20 | -.03 | .57 | .44 | .45 | .41 | 2.26 | 0.006 | 11.31 | 0.0000 | 28.40 | 0.0000 | 3.96 |
| 45 | S | .52 | .39 | .46 | .30 | -.39 | -.43 | -.29 | -.45 | -.16 | 1.11 | 0.344 | 4.31 | 0.0000 | 4.05 | 0.0002 | -0.15 |
| * 46 | S | .60 | .35 | .48 | .35 | -.26 | .49 | .35 | .31 | .42 | 2.02 | 0.017 | 7.64 | 0.0000 | 14.34 | 0.0000 | 2.62 |
| * 47 | S | .71 | .29 | .57 | .50 | -.19 | -.71 | -.21 | -.63 | -.04 | 3.10 | 0.000 | 12.94 | 0.0000 | 24.39 | 0.0000 | -0.18 |
| 48 | S | .60 | .42 | .47 | .38 | -.20 | -.61 | -.11 | -.51 | -.06 | 2.36 | 0.004 | 7.83 | 0.0000 | 8.22 | 0.0000 | 0.27 |

Die mit einem Stern versehenen Items wurden in die Fragebogen-Endform aufgenommen

.3.5.5. Itemselektion und Zusammenstellung der Fragebogenendform

Tabelle 13 a gibt einen Überblick über alle zur Itemselektion errechneten statistischen Kennwerte:

(1) Die Trennschärfe- und Schwierigkeitsindizes der Items bezogen auf die Subskalen Strenge (S) und Unterstützung (U) für die Gesamt- und Sonderschülerstichprobe;
(2) Die Korrelation zwischen den Items und dem Summenscore des Lügenfragebogens: ($r_{iL}$);
(3) Die F-Prüfgrößen einschließlich deren Signifikanzniveaus p' als Ergebnisse der Varianzanalysen zur Überprüfung der Itemmittelwertsunterschiede zwischen den Klassen getrennt nach Schultypen;
(4) Die Skalenwerte der Items als Ergebnis des Expertenratings.

Zunächst werden diejenigen Items ausgeschieden, in denen sich die Tendenz zu sozial erwünschtem Antwortverhalten manifestiert. Nach Tabelle 12 sind dies die Item-Nr.:

3. Er achtet auf Pünktlichkeit $r_{it} = 0,32$

6. Er hat es gern, wenn wir still auf unseren Plätzen sitzen $r_{it} = 0,40$

8. Der Lehrer erklärt schwierige Dinge und Aufgaben so, daß alle, auch die schwächeren Schüler der Klasse, es verstehen $r_{it} = 0,44$

11. Er achtet auf genaue Einhaltung von Pausenzeichen. $r_{it} = 0,45$

12. Wenn der Lehrer etwas erklärt hat, sagt er, daß es wohl nun auch dem Letzten klar geworden sein müsse. $r_{it} = 0,40$

17. Der Lehrer unterstützt Schüler, die in der Klasse nicht so beliebt sind. $r_{it} = 0,32$

22. Ich gehe gern zur Schule $r_{it} = 0,61$

45. An mir hat der Lehrer etwas auszusetzen. $r_{it} = 0,39$

Somit verbleibt noch ein Itempool von 40 Items mit Korrelationen von $r < 0,30$ zum Lügenfragebogen. Aus diesem Itempool werden folgende Items in die Endform des Schülerbeobachtungsbogens zur Lehrerbeurteilung aufgenommen:

(1) Items, die dem Expertenrating zufolge in extrem positivem und negativem Ausmaß zum Klassenklima und zur Lernatmosphäre beitragen, also Items um den Skalenwert 3 und 0 (vergl. Abbildung 2). Items aus dem Mittelbereich der Ratingskala bleiben unberücksichtigt; sie sind nach unserer Interpretation als Beitrag zum "Klassenklima und zur Lernatmosphäre" wenig trennwirksam.

(2) Items, für die zusätzlich zum Selektionskriterium (1) hohe Trennschärfen, bezogen auf die Subskalen Strenge und Unterstützung, errechnet wurden ($r_{it}$ sollte 0,40 nach Möglichkeit nicht unterschreiten).

(3) Items, die zusätzlich zu (1) und (2) hohe F-Werte als Kriterium, daß sie zwischen Lehrern differenzieren, aufweisen.

Berücksichtigen wir diesen Kriterienkatalog zur Itemselektion in der vorgegebenen Rangfolge, so verbleiben für die Endform die in Tabelle 13 b aufgeführten 25 Items, die, getrennt nach Gesamt- und Sonderschülerstichprobe, einer Itemanalyse unterzogen wurden. Die Reliabilitätsschätzungen (innere Konsistenz nach KUDER - RICHARDSON 21) betragen 0,90 für die Gesamtstichprobe und 0,84 für die Sonderschülerstichprobe.

Tab. 13 b

Endform des Fragebogens zur Erfassung des Interaktions- und Unterrichtsstils von Lehrern aus der Sicht der Schüler

| Item | Gesamtstichprobe | | Sonderschülerstichprobe | |
|---|---|---|---|---|
| | $r_{it}$ | P | $r_{it}$ | P |
| 1. Der Lehrer freut sich, wenn er die Klasse betritt. | .49 | .67 | .38 | .76 |
| 2. Wenn er etwas verspricht, dann hält er das auch. | .49 | .84 | .33 | .84 |
| 3. Der Lehrer hat schlechte Laune. | .57 | .54 | .52 | .55 |
| 4. Der Unterricht ist langweilig, daß man einschläft. | .57 | .63 | .54 | .67 |
| 5. Der Lehrer sagt, daß er Fragen von Schülern für dumm hält. | .41 | .74 | .37 | .76 |
| 6. Der Lehrer blamiert einige Schüler vor der Klasse. | .62 | .75 | .46 | .80 |
| 7. Der Lehrer sagt uns gegenüber manchmal Dinge, die wir zu ihm nicht sagen dürfen. | .62 | .69 | .42 | .77 |
| 8. Auch wenn der Lehrer bei Schülern Fehler entdeckt, bleibt er freundlich. | .57 | .85 | .46 | .81 |
| 9. Der Lehrer bespricht mit uns Schülern die Auswahl der Unterrichtsthemen. | .46 | .52 | .35 | .65 |
| 10. Der Lehrer erteilt Befehle. | .48 | .52 | .41 | .48 |
| 11. Der Lehrer gibt uns Aufgaben, ohne zu sagen, wie wir sie ausführen sollen. | .47 | .74 | .44 | .75 |
| 12. Der Lehrer lobt uns, wenn wir gute Arbeit leisten. | .38 | .77 | .23 | .81 |
| 13. Der Lehrer bringt uns Dinge bei, die wir gern können möchten. | .51 | .67 | .39 | .75 |
| 14. Wenn wir etwas getan haben, was der Lehrer falsch findet, dürfen wir sagen, warum wir es getan haben. | .59 | .73 | .48 | .69 |
| 15. Man kann mit ihm über alles sprechen, auch wenn man etwas ausgefressen hat. | .69 | .76 | .50 | .79 |
| 16. Der Lehrer sieht ein, daß die Schüler recht haben und gibt nach. | .61 | .65 | .51 | .60 |
| 17. Der Lehrer schimpft mit uns. | .57 | .40 | .45 | .42 |
| 18. Wenn in der Klasse Unruhe ist, dann bestraft der Lehrer die ganze Klasse. | .46 | .62 | .47 | .6[illegible] |
| 19. Wenn man ein Schulbuch vergißt, dann schimpft der Lehrer. | .57 | .54 | .53 | .56 |
| 20. Auch wenn man mal frech gewesen ist, bleibt der Lehrer ruhig. | .62 | .8[illegible] | .46 | .66 |
| 21. Der Lehrer sorgt dafür, daß alle Schüler arbeiten, so gut sie können. | .48 | .88 | .39 | .89 |
| 22. Der Lehrer gibt Strafarbeiten auf. | .57 | .49 | .51 | .[illegible]3 |
| 23. Bei unserem Lehrer dürfen wir offen unsere Meinung sagen. | .70 | .81 | .57 | .80 |
| 24. Der Lehrer ist auch zu den Schülern nett, die faul sind. | .60 | .64 | .48 | .65 |
| 25. Der Lehrer wird leicht ärgerlich. | .68 | .50 | .53 | .49 |

Mit der Endform wurde in drei Lernbehindertenklassen und einer Hauptschulklasse nach drei Monaten eine "Testwiederholung" vorgenommen. Diese "Testwiederholung" für den einen Hauptschullehrer erfolgte nicht mit derselben, sondern mit einer anderen Klasse. Korreliert wurden die Itemmittelwerte (N = 25 Meßwertpaare) aus der ersten und der zweiten Darbietung für jede der vier Klassen. Tabelle 13 c zeigt die Retest-Koeffizienten.

Tabelle 13 c

Retest-Koeffizienten der Endform

| Klasse (Schuljahr) | 7 (HS) | 7 (SS) | 8 (SS) | 9 (SS) |
|---|---|---|---|---|
| $r_{tt}$ | 0,96 | 0,77 | 0,76 | 0,93 |

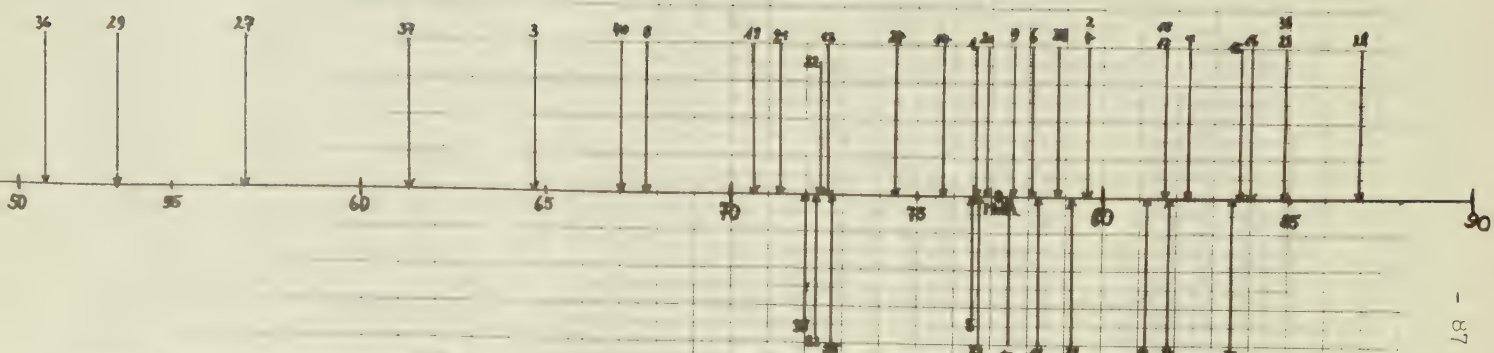



Abb. 3
Verteilung der Summenscores der Realurteile

# IV. ZWEITER TEIL DER HAUPTUNTERSUCHUNG

## 8. Validierungsaspekte der Fragebogenendform

### 8.1. Vorbemerkungen

In den vorangegangenen Kapiteln des ersten Hauptuntersuchungsteils wurden die einzelnen Stufen der Itemselektion und Konstruktion für die Endform des Fragebogens zur Erfassung des Lehrer-Erziehungsstils dargelegt. Dabei wurde das zweidimensionale Konzept Strenge und Unterstützung auf eine Dimension reduziert, die entsprechend der lernpsychologischen Terminologie "Bekräftigungswirkung des Lehrers" benannt wurde.

In diesem zweiten Hauptuntersuchungsteil werden einige Validierungsaspekte behandelt. Gesucht wird nach Lehrergruppen, die sich hinsichtlich ihrer Beurteilungsmodi (geschätzter, erwünschter und tatsächlich erhaltener Schülerurteile) und ihrer Diskrepanzen zwischen den Beurteilungsmodi unterscheiden. Hierbei handelt es sich um die Beantwortung der Fragestellungen 7 bis 12.

Die für die Endform ermittelten Summenscores der Real-, Schätz- und Wunschurteile der an der Untersuchung beteiligten Lehrer sind in Tabelle 14 aufgeführt. Lehrer-Nr. 17 und 18 beziehen sich auf denselben Lehrer, der von zwei verschiedenen siebten Klassen beurteilt wurde. Zusätzlich wird graphisch dargestellt, wie sich die Summenscores der Realurteile der an der Untersuchung beteiligten Lehrer verteilen (s. Abbildung 3).

In Abbildung 3 deutet sich eine rechtsgipflige Verteilung an. Rechtsgipfligkeit kann in diesem Falle durch einen Selektionseffekt bedingt sein. Es liegt nahe, anzunehmen,

Tab. 14

Summenscores der Real-, Schätz- und Wunschurteile der an der Untersuchung beteiligten Lehrer bezüglich der Fragebogen-Endform
( HS=Hauptschule; RS=Realschule; SS=Sonderschule)

| Lehrer Nr. | Schulart | Realurteile | Schätzurteile | Wunschurteile |
|---|---|---|---|---|
| 1 ♂ | SS | 76.6 | 87 | 94 |
| 2 ♀ | SS | 79.6 | 85 | 85 |
| 3 ♂ | SS | 64.7 | 58 | 83 |
| 4 ♂ | SS | 79.6 | 63 | 99 |
| 5 ♀ | SS | 76.5 | 66 | 81 |
| 6 ♀ | SS | 78.1 | 80 | 92 |
| 7 ♂ | SS | 82.3 | 67 | 94 |
| 8 ♂ | SS | 67.7 | 59 | 82 |
| 9 ♀ | SS | 77.6 | 74 | 79 |
| 10 ♀ | SS | 77.5 | 79 | 83 |
| 11 ♀ | SS | 78.3 | 76 | 91 |
| 12 ♂ | SS | 81.2 | 86 | 90 |
| 13 ♀ | SS | 72.6 | 77 | 91 |
| 14 ♂ | SS | 75.7 | 78 | 73 |
| 15 ♂ | SS | 83.7 | 76 | 87 |
| 16 ♂ | HS | 84.0 | 81 | 94 |
| 17 ♂ | HS | 81.7 | 89 | 94 |
| 18 ♂ | HS | 82.2 | 89 | 94 |
| 19 ♀ | HS | 70.6 | 74 | 85 |
| 20 ♀ | HS | 76.9 | 63 | 85 |
| 21 ♀ | HS | 71.3 | 80 | 92 |
| 22 ♀ | HS | 72.4 | 71 | 84 |
| 23 ♂ | HS | 84.9 | 79 | 84 |
| 24 ♀ | HS | 74.4 | 68 | 84 |
| 25 ♂ | HS | 72.7 | 74 | 99 |
| 26 ♀ | HS | 81.2 | 79 | 83 |
| 27 ♀ | HS | 56.9 | 78 | 80 |
| 28 ♂ | HS | 86.9 | 78 | 90 |
| 29 ♂ | HS | 52.6 | 63 | 72 |
| 30 ♂ | HS | 83.5 | 88 | 89 |
| 31 ♀ | RS | 79.2 | 69 | 84 |
| 32 ♂ | RS | 72.3 | 80 | 80 |
| 33 ♂ | RS | 76.7 | 83 | 87 |
| 34 ♂ | RS | 81.8 | 75 | 76 |
| 35 ♂ | RS | 72.0 | 72 | 91 |
| 36 ♂ | RS | 50.7 | 63 | 98 |
| 37 ♂ | RS | 61.3 | 70 | 79 |
| 38 ♂ | RS | 78.8 | 78 | 97 |
| 39 ♂ | RS | 84.9 | 75 | 88 |
| 40 ♀ | RS | 67.0 | 72 | 93 |

daß bei Lehrereinschätzungen durch Schüler bevorzugt erfolgreiche Lehrer teilnehmen, die sich ihren erfolgreichen Unterrichtsstil gern von ihren Schülern bescheinigen lassen. Der Vorteil einer rechtsgipfligen Verteilung kann andererseits darin erblickt werden, daß die Lehrer im unteren Bereich der Verteilung (also von ihren Schülern eher ungünstig eingestufte Lehrer) besser differenziert werden.

## 8.2. Statistische Auswertung des Zahlenmaterials im Hinblick auf die Validierungsfragen

Die Fragen 1 bis 6 in Kapitel 6 wurden bereits im Zusammenhang mit der Skalenkonstruktion und Itemselektion beantwortet. An dieser Stelle werden die Fragestellungen 7 bis 12 nacheinander aufgegriffen und bearbeitet.

Frage 7: Lassen sich bezüglich "angemessenen Interaktionsverhaltens im Unterricht" bei den an der Untersuchung beteiligten Lehrern schultypen- und/oder geschlechtsspezifische Unterschiede nachweisen?

Frage 7 wird dahingehend spezifiziert, ob Lehrer und Lehrerinnen in Sonderschulen und Lehrer und Lehrerinnen in Haupt- und Realschulen ein unterschiedlich angebrachtes Interaktionsverhalten aus der Sicht ihrer Schüler aufweisen. Diese Frage läßt sich mit einer zweifaktoriellen Varianzanalyse beantworten, in der Lehrergeschlecht und Schultyp die unabhängigen Variablen und das durch die Fragebogenendform gemessene Interaktionsverhalten die abhängige Variable darstellen. Somit ergibt sich folgendes Design:

| | Lehrer | | | Lehrerinnen | | |
|---|---|---|---|---|---|---|
| | Gruppe 1 | | | Gruppe 3 | | |
| Sonderschule | 76,58<br>64,66<br>79,64<br>67,66<br>77,57 | 81,23<br>75,69<br>83,72 | $\bar{X}$=75,84<br>N=8 | 79,64<br>76,47<br>78,10<br>82,26<br>77,46 | 78,29<br>72,56 | $\bar{X}$=77,82<br>N=7 |
| | Gruppe 2 | | | Gruppe 4 | | |
| Haupt- und Realschule | 84,03<br>81,68<br>84,86<br>72,72<br>86,95<br>52,63<br>83,45<br>72,32 | 76,10<br>81,77<br>72,00<br>50,74<br>61,28<br>78,81<br>84,93 | $\bar{X}$=74,94<br>N=15 | 70,58<br>76,92<br>71,31<br>72,41<br>74,38 | 81,20<br>56,90<br>79,18<br>67,00 | $\bar{X}$=72,20<br>N=9 |

Da ein multipler Mittelwertsvergleich zwischen 4 Gruppen interessiert, rechnen wir anstelle einer Varianzanalyse nach der Scheffé-Methode. Nach SCHEFFE (s. SACHS, 1969, S. 512) gilt für ungleich große und gleich große Stichprobengruppen

$$D_I = \sqrt{s^2_{in}\,(1/n_i + 1/n_j)(k-1)\,F_{(k-1;\,\nu^2_{sin};\,\alpha)}}$$

$s^2_{in}$ = Mittleres Quadrat des Versuchsfeldes
$n_i, n_j$ = Stichprobenumfänge der verglichenen Mittelwerte
$\nu^2_{sin}$ = Anzahl der Freiheitsgrade für $s^2_{in}$

Wird eine Irrtumswahrscheinlichkeit von $p \leq 0{,}05$ zugrundegelegt, so ergibt sich folgender multipler Vergleich:

| Indizes der Gruppen | | Differenz | Kritische Differenz |
|---|---|---|---|
| 1 | 2 | 0,90 | 11,32 |
| 1 | 3 | - 1,98 | 13,38 |
| 1 | 4 | 3,63 | 12,56 |
| 2 | 3 | - 2,88 | 11,83 |
| 2 | 4 | 2,73 | 10,90 |
| 3 | 4 | 5,61 | 13,03 |

Die kritische Differenz wird bei keinem Mittelwertsvergleich erreicht bzw. überschritten[1]. Die Nullhypothese, daß die vier Mittelwerte derselben Grundgesamtheit entstammen, kann beibehalten werden. Das Ausmaß angebrachten Interaktionsverhaltens ist - soweit es unsere Stichprobe betrifft - bei Lehrern und Lehrerinnen von Sonderschulen nicht von dem der Haupt- und Realschullehrer(innen) verschieden.

Frage 8: Unterscheiden sich besonders günstig beurteilte Lehrer im Vergleich zu eher ungünstig beurteilten Lehrern in ihren Schülerurteilen (vermuteten Realurteilen)?

Zunächst teilen wir die Lehrerstichprobe von N = 39 in zwei Hälften. Die Stichprobe reduziert sich von N = 40 auf N = 39, weil es sich bei Lehrer-Nr. 17 und 18 um denselben Lehrer handelt. Als cutt-off-Score wählen wir den Median (Mdn.), d.h. die Lehrerstichprobe wird nach den Ergebnissen der Fragebogenendform medianhalbiert, und wir erhalten dadurch eine obere (entsprechend den günstig beurteilten Lehrern) und eine untere Hälfte (entsprechend den ungünstig beurteilten Lehrern). Die zur oberen und unteren Hälfte gehörenden Lehrer werden danach unterschieden, ob sie mit ihren Schätzurteilen - bezogen auf den Summenscore der Fragebogenendform - über oder unter ihren Realurteilen liegen. Bei Lehrer-Nr. 35 sind die Summenscores der Real- und Schätzurteile identisch. Wir haben somit noch 38 Lehrer, die medianhalbiert werden. Den Mdn. erhalten wir durch das arithmetische Mittel aus den beiden mittelsten Werten, für die gleich viel größere und kleinere Werte existieren. Er beträgt 77,2 und ist in Abbildung 3 eingezeichnet.

---

1) Die Scheffé - Tests wurden auf einer PDP 8 - J mit einem Programm von HILLMANN gerechnet.

Die Frage, ob sich günstig und ungünstig beurteilte Lehrer in ihren Schätzurteilen voneinander unterscheiden, läßt sich mit dem $\chi^2$-Vierfelder-Test beantworten. Folgende Häufigkeitsangaben werden auf Gleichverteilung überprüft:

Summenscores der Realurteile

| | überschätzt | unterschätzt | |
|---|---|---|---|
| günstig beurteilt (>Mdn.) | 6 (a) | 13 (b) | 19 |
| ungünstig beurteilt (< Mdn.) | 13 (c) | 6 (d) | 19 |
| | 19 | 19 | 38 |

Setzen wir die Zahlen aus dieser Tafel in

$$\chi^2 = \frac{N\,(|ad - bc|)^2}{(a+b)\,(c+d)\,(a+c)\,(b+d)}$$ ein, so erhalten

wir ein $\chi^2$ von 5,15. Dieser Wert liegt bei 1 FG über der 5% - Signifikanzgrenze von 3,84.

Wir verwerfen die Nullhypothese, derzufolge über und unter dem Mdn. liegende Lehrer sich gleich häufig über- und unterschätzen. Die günstig beurteilte Lehrergruppe unter- und die ungünstig beurteilte Lehrergruppe überschätzt sich häufiger.

Frage 9: Weisen von ihren Schülern günstig eingestufte Lehrer im Vergleich zu eher ungünstig eingestuften Lehrern unterschiedliche Diskrepanzen zwischen
a) ihren Real- und Wunschurteilen und
b) ihren Wunsch- und Schätzurteilen auf?

Die Lehrerstichprobe von N = 39 wird wieder in zwei Hälften geteilt, in die obere und untere Gruppe. Nach

Tabelle 14 bilden wir für jeden Lehrer die Differenzen zwischen den Real- und Wunschurteilen und zwischen den Wunsch- und Schätzurteilen. Diese Differenzen tragen wir in Tabelle 15 ein.

Die mittlere Real- Wunschurteilsdifferenz und die mittlere Wunsch-Schätzurteilsdifferenz zwischen der oberen und unteren Lehrergruppe werden mit dem t-Test verglichen. Für ungleiche Stichprobenumfänge ( $n_1 \neq n_2$ ) wie in diesem Fall gilt:

$$t = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\left[\frac{n_1 + n_2}{n_1 \cdot n_2}\right]\left[\frac{Q_1 + Q_2}{n_1 + n_2 - 2}\right]}} \quad ; \quad FG = n_1 + n_2 - 2$$

$\bar{X}_1, \bar{X}_2$ = Mittelwerte der beiden Stichproben
$n_1, n_2$ = Stichprobenumfänge
$Q_1, Q_2$ = Summe der Abweichungsquadrate.

Der Rechengang für beide t-Tests erfolgte wieder auf der PDP 8-J nach einem Programm von HIMMELMANN, das die Prüfung auf Varianzhomogenität einschließt.
Die Differenz von 9,6 zwischen den mittleren Real-Wunschurteilsdifferenzen der beiden Lehrergruppen ist hochsignifikant: t = 3,49; df = 37; $P < 0{,}002$.

Die Differenz von 3,17 zwischen den mittleren Wunsch-Schätzurteilsdifferenzen der beiden Lehrergruppen ist nicht signifikant: t = 1,03; df = 37.

Günstig beurteilte Lehrer weisen geringere Diskrepanzen zwischen den Real- und Wunschurteilen auf als ungünstig beurteilte Lehrer.

Frage 10: Lassen sich aufgrund der HOFERSCHEN Beurteilungsdimensionen zur aktuellen Schülerbeurteilung unterschiedliche Lehrerbeurteilertypen finden?

Tab. 15

Differenzen zwischen den Real- und Wunschurteilen und zwischen den Wunsch- und Schätzurteilen für die nach der Fragebogen-Endform günstig und ungünstig eingestufte Lehrergruppe*

| Lehrer Nr. | Differenzen zwischen den Real- und Wunschurteilen | Differenzen zwischen den Wunsch- und Schätzurteilen | |
|---|---|---|---|
| 1 | 19.2 | 12 | |
| 3 | 24.1 | 30 | |
| 5 | 10.3 | 20 | |
| 8 | 25.2 | 28 | ungünstig eingestufte Lehrergruppe (in der Schülerbewertung unter dem Median liegend) |
| 13 | 24.2 | 19 | |
| 14 | 3.1 | 0 | |
| 19 | 20.2 | 16 | |
| 20 | 13.9 | 27 | |
| 21 | 26.5 | 17 | |
| 22 | 17.4 | 18 | |
| 24 | 15.4 | 21 | |
| 25 | 32.1 | 30 | |
| 27 | 28.9 | 7 | |
| 29 | 25.2 | 14 | |
| 32 | 13.5 | 5 | |
| 33 | 16.1 | 9 | |
| 35 | 24.8 | 24 | |
| 36 | 53.1 | 40 | |
| 37 | 23.5 | 14 | |
| 40 | 31.8 | 26 | |
| 2 | 11.2 | 5 | |
| 4 | 25.2 | 41 | |
| 6 | 19.7 | 17 | |
| 7 | 17.5 | 32 | |
| 9 | 7.2 | 10 | |
| 10 | 11.3 | 9 | |
| 11 | 18.5 | 20 | |
| 12 | 14.6 | 9 | |
| 15 | 9.1 | 16 | |
| 16 | 15.8 | 18 | |
| 17 | 18.1 | 10 | |
| 23 | 4.9 | 10 | |
| 26 | 7.6 | 9 | günstig eingestufte Lehrergruppe ( in der Schülerbewertung über dem Median liegend) |
| 28 | 8.9 | 18 | |
| 30 | 11.3 | 6 | |
| 31 | 10.6 | 20 | |
| 34 | 0.0 | 6 | |
| 38 | 24.0 | 24 | |
| 39 | 8.9 | 18 | |

* Da sich positive Differenzen rechnerisch besser verarbeiten lassen, wurden zu den Differenzen entsprechende additive Konstanten hinzugefügt.

Bevor diese Frage durch Clusteranalysen bearbeitet wird, sollen die HOFERSCHEN Polaritäten (s. Anhang, S. 143 ), anhand derer die Lehrer jeden ihrer Schüler zu beurteilen hatten, faktoren- und itemanalysiert werden [1]. Da drei Lehrer (nach Tabelle 14 Lehrer-Nr. 24, 25 und 37) keine Schülerbeurteilungen mit den Polaritäten durchführten, reduziert sich die Gesamtstichprobe von N = 937 um drei Klassen auf N = 810 beurteilte Schüler.

In die Faktorenanalyse (R - Analyse) mit anschließender Varimax-Rotation gehen die 25 Polaritäten des Schülerbeurteilungsbogens als Variablen und die 810 beurteilten Schüler als Meßwertträger ein. Als Ergebnis dieser R -Analyse können die 25 polaren Eigenschaften zu den in Tabelle 16 aufgeführten Faktoren zusammengefaßt werden. Tabelle 16 enthält außerdem die Trennschärfekoeffizienten $r_{it}$ der einzelnen polaren Eigenschaften (Items) bezogen auf den jeweiligen als Subtest betrachteten Faktor, zudem die Mittelwerte $\bar{x}$ der Polaritäten. Bis auf Faktor IV, den wir als "Normenkonformität" interpretierten, konnte die Benennung der weiteren Faktoren nach HOFER (1969, S. 60) erfolgen.

Der Reliabilitätskoeffizient der Dimension Empfindsamkeit vs. Robustheit fällt gegenüber den anderen Dimensionen auf 0,52 ab, da dieser "Subtest" nur aus zwei Items (Polaritäten) besteht. Die übrigen Beurteilungsdimensionen können, angesichts ihres geringen Itemumfangs, als hoch reliabel angesehen werden.

Tabelle 17 a gibt die durchschnittlichen Beurteilungsscores eines jeden Lehrers für die Schüler seiner Klasse auf den fünf Beurteilungsdimensionen wieder. Der durchschnittliche Beurteilungsscore pro Lehrer auf einer

---

1) Der Zweck der Faktorenanalyse an dieser Stelle ist es, die 25 polaren Eigenschaften (Items) zu Subtests zusammenzufassen.

Tab. 16

Ergebnisse der Faktoren- und Itemanalyse der polaren Eigenschaften zur aktuellen Schülerbeurteilung

| | | Faktorenladungen | | | | | | |
|---|---|---|---|---|---|---|---|---|
| I. | Disziplniertheit vs. Undisziplniertheit | I | II | III | IV | V | $r_{it}$ | $\bar{x}$ |
| | 1. uninteressiert/interessiert | -.56 | -.27 | -.51 | -.20 | .03 | .71 | 3.02 |
| | 9. fleißig/faul | .83 | -.06 | .14 | .22 | .03 | .85 | 3.22 |
| | 12. ehrgeizig/gleichgültig | .83 | .17 | .21 | .09 | .14 | .81 | 3.63 |
| | 13. unkonzentriert/konzentriert | -.71 | .24 | -.33 | -.05 | .24 | .81 | 3.80 |
| | 15. unordentlich/ordentlich | -.60 | .23 | -.13 | -.27 | .08 | .71 | 3.21 |
| | 19. pflichtbewußt/ pflichtvergessen | .83 | -.12 | .11 | .24 | -.02 | .85 | 3.11 |
| | 20. aufmerksam/abgelenkt | .75 | .18 | .32 | .15 | -.18 | .85 | 3.55 |
| | 23. zuverlässig/ unzuverlässig | .72 | -.06 | .16 | .41 | -.10 | .83 | 2.94 |
| II. | Zurückhaltung vs Geltungsstreben | | | | | | | |
| | 2. verschlossen/ offen | -.19 | -.72 | -.15 | -.13 | -.28 | .72 | 3.46 |
| | 4. selbstsicher/ selbstunsicher | .05 | .49 | .42 | -.01 | -.47 | .67 | 3.81 |
| | 7. gesellig/ ungesellig | -.03 | .68 | .21 | .33 | -.22 | .68 | 2.97 |
| | 8. geltungsbedürftig/ bescheiden | -.25 | .68 | .04 | -.38 | .07 | .74 | 4.19 |
| | 16. schüchtern/ dreist | -.24 | -.71 | -.09 | .41 | .12 | .79 | 4.17 |
| | 18. unruhig/ ruhig | -.39 | .73 | .03 | -.15 | .15 | .74 | 4.62 |
| | 25. führend/ zurückhaltend | -.04 | .75 | .20 | -.15 | -.18 | .82 | 4.19 |
| III. | Geistige Regsamkeit | | | | | | | |
| | 5. unbegabt/ begabt | -.23 | -.01 | -.89 | -.08 | .04 | .88 | 3.41 |
| | 6. einfallsreich/ einfallslos | .22 | .33 | .71 | .05 | -.05 | .83 | 3.64 |
| | 11. intelligent/ dumm | .21 | .07 | .87 | .08 | .05 | .88 | 3.40 |
| | 21. aufgeweckt/träge | .34 | .43 | .62 | .09 | -.07 | .83 | 3.37 |
| IV. | Normenkonformität | | | | | | | |
| | 14. höflich/ frech | .31 | -.19 | .06 | .72 | .01 | .79 | 2.58 |
| | 17. folgsam/ unfolgsam | .40 | -.33 | -.03 | .65 | -.11 | .84 | 2.84 |
| | 22. ausgeglichen/ launisch | .33 | -.28 | .09 | .53 | -.44 | .80 | 3.29 |
| | 24. sympathisch/ unsympathisch | .21 | .21 | .24 | .70 | .00 | .67 | 2.65 |
| V. | Empfindsamkeit vs. Robustheit | | | | | | | |
| | 3. sensibel/ robust | .17 | -.44 | .00 | .26 | .65 | .83 | 3.69 |
| | 10. kompliziert/ unkompliziert | -.20 | -.18 | .05 | -.24 | .74 | .81 | 4.14 |

Die Reliabilitätsschätzungen nach KUDER RICHARDSON 21 für die einzelnen als Subtests betrachteten Dimensionen fallen wie folgt aus:

| | | |
|---|---|---|
| I. | Disziplniertheit vs. Undisziplniertheit | .92 |
| II. | Zurückhaltung vs. Geltungsstreben | .86 |
| III. | Geistige Regsamkeit | .87 |
| IV. | Normenkonformität | .78 |
| V. | Empfindsamkeit vs. Robustheit | .52 |

Tab. 17 a

Durchschnittliche Beurteilungsscores der Lehrer bezüglich der fünf Beurteilungsdimensionen zur aktuellen Schülerbewertung *

| Lehrer Nr. | I | $T_{(I)}$ | II | $T_{(II)}$ | III | $T_{(III)}$ | IV | $T_{(IV)}$ | V | $T_{(V)}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 25.5 | 47.3 | 21.0 | 26.1 | 12.3 | 42.8 | 10.1 | 43.7 | 6.5 | 40.1 |
| 2 | 21.2 | 35.7 | 21.3 | 27.3 | 10.2 | 33.3 | 7.2 | 28.5 | 8.7 | 56.9 |
| 3 | 24.6 | 44.9 | 28.4 | 54.6 | 11.8 | 40.5 | 14.3 | 65.7 | 7.3 | 46.2 |
| 4 | 34.4 | 71.3 | 28.2 | 53.1 | 18.6 | 71.2 | 15.2 | 70.4 | 4.6 | 25.6 |
| 5 | 26.6 | 50.3 | 23.5 | 35.8 | 17.9 | 68.0 | 10.8 | 47.6 | 8.1 | 47.7 |
| 6 | 33.1 | 67.6 | 28.3 | 54.2 | 16.9 | 63.5 | 10.9 | 47.9 | 6.2 | 47.8 |
| 7 | 26.7 | 50.5 | 27.8 | 51.9 | 16.3 | 60.8 | 11.7 | 52.1 | 7.3 | 46.2 |
| 8 | 32.6 | 66.4 | 31.2 | 65.4 | 15.1 | 55.4 | 11.5 | 51.1 | 6.4 | 49.3 |
| 9 | 27.2 | 51.9 | 30.2 | 61.5 | 15.0 | 55.0 | 11.1 | 48.9 | 8.6 | 56.1 |
| 10 | 18.7 | 29.0 | 25.9 | 45.0 | 11.9 | 41.0 | 9.2 | 39.0 | 9.9 | 66.0 |
| 11 | 21.9 | 37.6 | 26.3 | 46.5 | 9.3 | 29.3 | 9.3 | 39.5 | 4.1 | 21.8 |
| 12 | 23.7 | 42.4 | 28.5 | 55.0 | 13.0 | 45.9 | 10.1 | 43.7 | 7.8 | 50.0 |
| 13 | 27.7 | 53.2 | 27.3 | 50.4 | 14.3 | 51.8 | 10.1 | 43.7 | 7.6 | 48.5 |
| 14 | 25.8 | 48.1 | 24.5 | 39.6 | 13.7 | 49.1 | 10.4 | 45.3 | 10.7 | 72.1 |
| 15 | 28.8 | 56.2 | 28.4 | 54.6 | 17.4 | 65.8 | 13.4 | 61.0 | 7.8 | 50.0 |
| 16 | 28.9 | 56.5 | 29.3 | 58.1 | 15.8 | 58.6 | 13.4 | 61.0 | 7.5 | 47.7 |
| 17 | 21.5 | 36.5 | 26.4 | 46.9 | 10.4 | 34.2 | 9.9. | 42.7 | 9.6 | 63.7 |
| 18 | - | - | - | - | - | - | - | - | - | - |
| 19 | 22.9 | 40.3 | 24.3 | 38.8 | 11.4 | 38.7 | 13.0 | 58.9 | 8.9 | 58.4 |
| 20 | 31.0 | 62.1 | 26.1 | 45.8 | 14.2 | 51.4 | 11.3 | 50.0 | 7.9 | 50.8 |
| 21 | 27.4 | 52.4 | 29.7 | 59.6 | 13.4 | 47.7 | 10.6 | 46.3 | 8.9 | 58.4 |
| 22 | 25.2 | 46.5 | 30.7 | 63.5 | 13.1 | 46.4 | 11.3 | 50.0 | 7.6 | 48.5 |
| 23 | 28.9 | 56.5 | 25.7 | 44.2 | 14.1 | 50.9 | 10.8 | 47.4 | 7.0 | 43.9 |
| 24 | - | - | - | - | - | - | - | - | - | - |
| 25 | - | - | - | - | - | - | - | - | - | - |
| 26 | 32.5 | 66.2 | 28.3 | 54.2 | 17.3 | 65.3 | 13.5 | 62.0 | 7.8 | 50.0 |
| 27 | 24.2 | 43.8 | 24.2 | 38.5 | 12.5 | 43.7 | 7.6 | 30.6 | 8.7 | 56.9 |
| 28 | 28.1 | 54.3 | 26.7 | 48.1 | 14.9 | 54.5 | 10.2 | 44.2 | 8.9 | 58.4 |
| 29 | 20.7 | 34.4 | 29.0 | 56.9 | 12.0 | 41.4 | 9.1 | 38.5 | 6.4 | 39.3 |
| 30 | 27.7 | 53.2 | 29.7 | 59.6 | 15.2 | 55.9 | 10.5 | 45.8 | 9.4 | 62.2 |
| 31 | 24.1 | 43.5 | 27.7 | 51.9 | 12.4 | 43.2 | 9.9 | 57.3 | 7.5 | 47.7 |
| 32 | 23.1 | 40.8 | 28.8 | 56.2 | 12.9 | 45.5 | 13.4 | 61.0 | 8.5 | 55.3 |
| 33 | 25.4 | 47.0 | 28.7 | 55.8 | 12.2 | 42.3 | 10.2 | 44.2 | 6.8 | 42.4 |
| 34 | 29.3 | 57.6 | 28.0 | 53.1 | 14.6 | 53.2 | 14.7 | 67.8 | 7.3 | 46.2 |
| 35 | 23.5 | 40.9 | 24.7 | 40.4 | 11.6 | 39.6 | 13.8 | 63.1 | 7.8 | 50.0 |
| 36 | 28.5 | 55.4 | 24.6 | 40.0 | 13.5 | 48.2 | 12.7 | 57.3 | 7.9 | 50.8 |
| 37 | - | - | - | - | - | - | - | - | - | - |
| 38 | 29.9 | 59.2 | 24.3 | 38.8 | 14.8 | 54.1 | 11.6 | 51.6 | 8.9 | 58.4 |
| 39 | 29.4 | 57.3 | 30.4 | 62.3 | 16.2 | 60.4 | 13.3 | 60.5 | 8.1 | 52.3 |
| 40 | 23.1 | 40.8 | 32.0 | 68.5 | 15.4 | 56.8 | 8.9 | 37.4 | 8.4 | 54.6 |

* Zum Zweck späterer Profildarstellung sind die durchschnittlichen Beurteilungsscores in T - Normen überführt worden.

Dimension wird ermittelt, indem die Summe aus allen zu einer Dimension gehörenden polaren Eigenschaften (Items) für alle Schüler einer Klasse durch die Anzahl der Schüler der jeweiligen Klassen dividiert wird. Ein hoher durchschnittlicher Punktwert für einen Lehrer auf beispielsweise der Dimension geistige Regsamkeit bedeutet, daß der entsprechende Lehrer dazu neigt, seine Schüler in bezug auf "geistige Regsamkeit" streng zu bewerten, ein niedriger Punktwert bedeutet umgekehrt, daß der entsprechende Lehrer eher eine mildere Bewertung für seine Schüler vornimmt. Aus Tabelle 17 a wird ersichtlich, daß die durchschnittlichen Beurteilungsscores der Lehrer hinsichtlich der fünf Beurteilungsdimensionen beträchtlich variieren; das bedeutet, daß die Lehrer nach diesen Beurteilungsdimensionen zu unterschiedlich strengen Schülerurteilen neigen, ihr implizites Persönlichkeitsmodell zur aktuellen Schülerbeurteilung gewissermaßen mehr oder weniger streng realisieren.

Die Frage, ob sich verschiedene, abgrenzbare Lehrerbeurteilertypen unterscheiden lassen, bearbeiten wir mit der Clusteranalyse als Methode statistischer Typenfindung. Wir suchen nach unterscheidbaren Lehrergruppen, die aufgrund ihrer Ähnlichkeit bezüglich der durchschnittlichen Beurteilungsscores auf den fünf Urteilsdimensionen zusammengefaßt sind.

Auch die Faktorenanalyse stellt eine Methode zur statistischen Typenfindung dar. Während in der Clusteranalyse Gruppen ähnlicher Meßwertträger gesucht werden, arbeitet die Faktorenanalyse (R-Analyse) zugrundeliegende Dimensionen (Typusfaktoren) heraus. Dabei können zum Teil unähnliche Meßwertträger (Elemente) im selben Faktor laden. Allerdings sagt gleiche Ladungshöhe im selben Faktor noch nichts über die Ähnlichkeit der Meßwertträger aus, weil Unterschiede hinsichtlich der Lage (evaluation) zweier Profile

und die Varianzunterschiede zweier Profile (scatter) nicht in die Faktorenanalyse eingehen. Ein Verfahren, welches diese beiden Profilaspekte berücksichtigt, findet sich bei VELDMAN (1968, S. 308) in der Hierarchical Grouping Analysis nach WARD (1963), die mit dem Programm HGROUP automatisiert worden ist. Das Verfahren betrachtet zunächst die N Profile (bezüglich K Variabler) der zu Gruppen zusammenzufassenden Meßwertträger als N Gruppen. Die beiden ähnlichsten dieser N Gruppen werden zunächst zusammengefaßt, es resultieren N - 1 Gruppen. Das Kriterium für die Zusammenfassung zweier dieser N Gruppen bildet die relativierte Quadratsumme der Abweichungen dieser beiden Gruppen von ihrem gemeinsamen Mittelwertsprofil. Dieses Kriterium zeigt gleichzeitig das Ausmaß des Fehlerzuwachses an, den man bei iterierender Zusammenfassung von Gruppen begeht. Nun wird bei den N - 1 Gruppen dasselbe Verfahren wiederholt. Als Abbruchkriterium dient hierbei die Inspektion des Fehlerzuwachses nach jeder Iteration. Man entscheidet sich für diejenige Gruppierung, die - verglichen mit vorherigen Iterationen - von einem sprunghaften Fehlerzuwachs liegt.

Mit der Hierarchical Group Analysis ermitteln wir, ob sich aus den Daten der Tabelle 17 a Typen (Gruppen) unterschiedlich urteilender Lehrer finden lassen. Da 36 Lehrer zu gruppieren sind, gehen in die Clusteranalyse 36 Profile, beruhend auf fünf Beurteilungsdimensionen (Variablen), ein. Nach dem Kriterium des sprunghaften Fehlerzuwachses entscheiden wir uns, wie aus Tabelle 17 b hervorgeht, für vier Lehrergruppen, deren durchschnittliche Beurteilungsscores bezüglich der fünf Urteilsdimensionen in Tabelle 17 b mit aufgeführt sind.

Um zu ermitteln, welche Beurteilungsdimensionen sich für die vorliegende Gruppierung als trennwirksam erwiesen

Tab. 17 b

Lehrergruppen unterschiedlicher Urteilstendenzen bezüglich der fünf Beurteilungsdimensionen nach der 1. Clusteranalyse

| Lehrer Nr. | I | II | III | IV | V | |
|---|---|---|---|---|---|---|
| 1 | 25.5 | 21.0 | 12.3 | 10.1 | 6.5 | |
| 2 | 21.2 | 21.3 | 12.2 | 7.2 | 8.7 | |
| 10 | 18.7 | 25.9 | 11.9 | 9.2 | 9.9 | |
| 14 | 25.8 | 24.5 | 13.7 | 10.4 | 10.7 | Gruppe 1 |
| 17 | 21.5 | 26.4 | 10.4 | 9.9 | 9.6 | |
| 27 | 24.2 | 24.2 | 12.5 | 7.6 | 8.7 | |
| $\bar{X}$ | 22.8 | 23.9 | 11.8 | 9.0 | 9.0 | |
| 3 | 24.6 | 28.4 | 11.8 | 14.3 | 7.3 | |
| 11 | 21.9 | 26.3 | 9.3 | 9.3 | 4.1 | |
| 12 | 23.7 | 28.5 | 13.0 | 10.1 | 7.8 | |
| 19 | 22.9 | 24.3 | 11.4 | 13.0 | 8.9 | |
| 22 | 25.2 | 30.7 | 13.1 | 11.3 | 7.6 | Gruppe 2 |
| 29 | 20.7 | 29.0 | 12.0 | 9.1 | 6.4 | |
| 31 | 24.1 | 27.7 | 12.4 | 9.9 | 7.5 | |
| 32 | 23.1 | 28.8 | 12.9 | 13.4 | 8.5 | |
| 33 | 25.4 | 28.7 | 12.2 | 10.2 | 6.8 | |
| 35 | 23.5 | 24.7 | 11.6 | 13.8 | 7.8 | |
| $\bar{X}$ | 23.5 | 27.7 | 12.0 | 11.4 | 7.3 | |
| 4 | 34.4 | 28.2 | 18.6 | 15.2 | 4.6 | |
| 6 | 33.1 | 28.3 | 16.9 | 10.9 | 6.2 | |
| 8 | 32.6 | 31.2 | 15.1 | 11.5 | 6.4 | |
| 15 | 28.8 | 28.4 | 17.4 | 13.4 | 7.8 | Gruppe 3 |
| 16 | 28.9 | 29.3 | 15.8 | 13.4 | 7.5 | |
| 26 | 32.5 | 28.3 | 17.3 | 13.5 | 7.8 | |
| 34 | 29.3 | 28.0 | 14.6 | 14.7 | 7.3 | |
| 39 | 29.2 | 30.4 | 16.2 | 13.3 | 8.1 | |
| $\bar{X}$ | 31.1 | 29.0 | 16.5 | 13.2 | 7.0 | |
| 5 | 26.6 | 23.5 | 17.9 | 10.8 | 8.1 | |
| 7 | 26.7 | 27.8 | 16.3 | 11.7 | 7.3 | |
| 9 | 27.2 | 30.2 | 15.0 | 11.1 | 8.6 | |
| 13 | 27.7 | 27.3 | 14.3 | 10.1 | 7.6 | |
| 20 | 31.0 | 26.1 | 14.2 | 11.3 | 7.9 | |
| 21 | 27.4 | 29.7 | 13.4 | 10.6 | 8.9 | |
| 23 | 28.9 | 25.7 | 14.1 | 10.8 | 7.0 | Gruppe 4 |
| 28 | 28.1 | 26.7 | 14.9 | 10.2 | 8.9 | |
| 30 | 27.7 | 29.7 | 15.2 | 10.5 | 9.4 | |
| 36 | 28.5 | 24.6 | 13.5 | 12.7 | 7.9 | |
| 38 | 29.9 | 24.3 | 14.8 | 11.6 | 8.9 | |
| 40 | 23.1 | 32.0 | 15.4 | 8.9 | 8.4 | |
| $\bar{X}$ | 27.7 | 25.2 | 14.9 | 10.9 | 8.2 | |
| $\bar{X}_{(gesamt)}$ | 26.5 | 27.4 | 13.8 | 11.4 | 7.8 | |
| s | 3.7 | 2.6 | 2.2 | 1.9 | 1.3 | |

haben, prüfen wir die Mittelwertsunterschiede zwischen den vier Gruppen bezüglich jeder Beurteilungsdimension auf Signifikanz. Die multiplen Mittelwertvergleiche werden wieder mit Hilfe des SCHEFFE - Tests vorgenommen (vergl. SACHS, a.a.O.). Wird eine Irrtumswahrscheinlichkeit von $p \leqslant 0{,}05$ zugrundegelegt, so ergeben sich die folgenden multiplen Vergleiche:

I) Bezüglich der Dimension Disziplinertheit vs. Undisziplinertheit:

| Indizes der Gruppen | | Differenz | Kritische Differenz |
|---|---|---|---|
| 1 | 2 | -0,66 | 3,15 |
| * 1 | 3 | -8,28 | 3,29 |
| * 1 | 4 | -4,91 | 3,05 |
| * 2 | 3 | -7,61 | 2,89 |
| * 2 | 4 | -4,25 | 2,61 |
| * 3 | 4 | 3,36 | 2,78 |

II) Bezüglich der Dimension Zurückhaltung vs. Geltungsstreben:

| Indizes der Gruppen | | Differenz | Kritische Differenz |
|---|---|---|---|
| 1 | 2 | -3,85 | 7,66 |
| 1 | 3 | -5,16 | 8,02 |
| 1 | 4 | -1,34 | 7,33 |
| 2 | 3 | -1,30 | 7,04 |
| 2 | 4 | 2,51 | 6,24 |
| 3 | 4 | 3,82 | 6,67 |

III) Bezüglich der Dimension geistige Regsamkeit:

| Indizes der Gruppen | | Differenz | Kritische Differenz |
|---|---|---|---|
| 1 | 2 | - 0,12 | 1,88 |
| * 1 | 3 | -4,65 | 1,97 |
| * 1 | 4 | -3,08 | 1,82 |
| * 2 | 3 | -4,25 | 1,73 |
| * 2 | 4 | -2,95 | 1,56 |
| 3 | 4 | 1,57 | 1,66 |

IV) Bezüglich der Dimension Normenkonformität:

| Indizes der Gruppen | | Differenz | Kritische Differenz |
|---|---|---|---|
| * 1 | 2 | -2,37 | 2,24 |
| * 1 | 3 | -4,17 | 2,34 |
| 1 | 4 | -1,79 | 2,17 |
| 2 | 3 | -1,79 | 2,06 |
| 2 | 4 | 0,58 | 1,86 |
| * 3 | 4 | 2,37 | 1,98 |

V) Bezüglich der Dimension Empfindsamkeit vs. Robustheit:

| Indizes der Gruppen | | Differenz | Kritische Differenz |
|---|---|---|---|
| * 1 | 2 | 1,74 | 1,74 |
| * 1 | 3 | 2,05 | 1,82 |
| 1 | 4 | 0,77 | 1,68 |
| 2 | 3 | 0,30 | 1,60 |
| 2 | 4 | -0,97 | 1,44 |
| 3 | 4 | -1,27 | 1,54 |

Die signifikanten Mittelwertsdifferenzen sind mit einem Stern versehen. Wir erkennen, daß sich für die Dimension Zurückhaltung vs. Geltungsstreben keine signifikanten Differenzen ergeben haben und betrachten daher diese Dimension für die Gruppentrennung als unbedeutsam. Deshalb rechneten wir noch eine zweite Clusteranalyse, in die die Dimension Zurückhaltung vs. Geltungsstreben nicht mehr einging. Das Ergebnis dieser zweiten Clusteranalyse wird durch Tabelle 18 ausgewiesen. Ohne die Variable Zurückhaltung vs. Geltungsstreben erhalten wir also jetzt eine Dreier-Gruppierung als beste Lösung.

Die Mittelwertsunterschiede zwischen diesen drei erhaltenen Gruppen werden wieder mit dem SCHEFFE - Test auf Signifikanz geprüft. Bei einer Irrtumswahrscheinlichkeit von $p \leq 0{,}05$ erhalten wir folgende multiple Vergleiche:

Tab. 18

Lehrergruppen unterschiedlicher Urteilstendenzen als Ergebnis der 2. Clusteranalyse

| Lehrer Nr. | I | III | IV | V | |
|---|---|---|---|---|---|
| 1 | 25.5 | 12.3 | 10.1 | 6.5 | |
| 2 | 21.2 | 10.2 | 7.2 | 8.7 | |
| 10 | 18.7 | 11.9 | 9.2 | 9.9 | |
| 11 | 21.6 | 9.3 | 9.3 | 4.9 | |
| 12 | 23.7 | 13.0 | 10.1 | 7.8 | Gruppe 1 |
| 17 | 21.5 | 10.4 | 9.9 | 9.6 | |
| 27 | 24.2 | 12.5 | 7.6 | 8.7 | |
| 29 | 20.7 | 11.9 | 9.1 | 6.4 | |
| 31 | 24.1 | 12.4 | 9.1 | 7.5 | |
| 33 | 25.4 | 12.2 | 10.2 | 6.8 | |
| $\bar{X}$ | 22.7 | 11.6 | 9.2 | 7.7 | |
| $\bar{T}$ | 39.8 | 39.6 | 39.0 | 49.3 | |
| 3 | 24.6 | 11.8 | 14.3 | 7.3 | |
| 5 | 26.6 | 17.9 | 10.8 | 8.1 | |
| 7 | 26.7 | 16.3 | 11.7 | 7.3 | |
| 9 | 27.2 | 15.0 | 11.1 | 8.6 | |
| 13 | 27.7 | 14.3 | 10.1 | 7.6 | |
| 14 | 25.8 | 13.7 | 10.4 | 10.7 | |
| 19 | 22.9 | 11.4 | 13.0 | 8.9 | |
| 20 | 31.0 | 14.2 | 11.3 | 7.9 | |
| 21 | 27.4 | 13.4 | 10.6 | 8.9 | |
| 22 | 25.2 | 13.1 | 11.8 | 7.6 | Gruppe 2 |
| 23 | 28.9 | 14.1 | 10.8 | 7.0 | |
| 28 | 28.1 | 14.9 | 10.2 | 8.9 | |
| 30 | 27.7 | 15.2 | 10.5 | 9.4 | |
| 32 | 23.1 | 12.9 | 13.4 | 8.4 | |
| 35 | 23.5 | 11.6 | 13.8 | 7.8 | |
| 36 | 28.5 | 13.5 | 12.7 | 7.9 | |
| 38 | 29.9 | 14.8 | 11.6 | 8.9 | |
| 40 | 23.1 | 15.4 | 8.9 | 8.4 | |
| $\bar{X}$ | 26.5 | 14.1 | 11.5 | 8.3 | |
| $\bar{T}$ | 50.2 | 50.9 | 51.1 | 53.4 | |
| 4 | 34.4 | 18.6 | 15.2 | 4.6 | |
| 6 | 33.1 | 16.9 | 10.9 | 6.2 | |
| 8 | 32.6 | 15.1 | 11.5 | 6.4 | |
| 15 | 28.8 | 17.4 | 13.4 | 7.8 | |
| 16 | 28.9 | 15.8 | 13.4 | 7.5 | Gruppe 3 |
| 26 | 32.5 | 17.3 | 13.5 | 7.8 | |
| 34 | 29.3 | 14.6 | 14.7 | 7.3 | |
| 39 | 29.2 | 16.2 | 13.3 | 8.1 | |
| $\bar{X}$ | 31.1 | 16.5 | 13.2 | 6.9 | |
| T | 62.3 | 61.3 | 60.2 | 42.6 | |

I) Bezüglich der Dimension Diszipliniertheit vs. Undiszipliniertheit:

| Indizes der Gruppen | | Differenz | Kritische Differenz |
|---|---|---|---|
| * 1 | 2 | -3,88 | 2,36 |
| * 1 | 3 | -8,44 | 2,84 |
| * 2 | 3 | -4,55 | 2,54 |

III) Bezüglich der Dimension geistige Regsamkeit:

| Indizes der Gruppen | | Differenz | Kritische Differenz |
|---|---|---|---|
| * 1 | 2 | -2,47 | 1,48 |
| * 1 | 3 | -4,87 | 1,79 |
| * 2 | 3 | -2,40 | 1,60 |

IV) Bezüglich der Dimension Normenkonformität:

| Indizes der Gruppen | | Differenz | Kritische Differenz |
|---|---|---|---|
| * 1 | 2 | -2,31 | 1,35 |
| * 1 | 3 | -4,05 | 1,63 |
| * 2 | 3 | -1,73 | 1,46 |

V) Bezüglich der Dimension Empfindsamkeit vs. Robustheit:

| Indizes der Gruppen | | Differenz | Kritische Differenz |
|---|---|---|---|
| 1 | 2 | -0,63 | 1,18 |
| 1 | 3 | 0,71 | 1,42 |
| * 2 | 3 | 1,35 | 1,28 |

Die Bezeichnung (II) "römisch zwei" entfällt, da die Dimension Zurückhaltung vs. Geltungsstreben in die 2. Clusteranalyse nicht eingeht.

Der Ausdruck des Rechenprotokolls (Tabelle 19) der 2. Clusteranalyse läßt erkennen, daß nach der Dreier-Gruppierung ein sprunghafter Fehleranstieg von 14,7270 auf 25,4169 zu verzeichnen ist.

Tab. 19 Ausdruck des Rechenprotokolls der 2. Clusteranalyse

```
35 GRUPPEN NACH VEREINIGUNG VON GRUPPE   1 (N =  1) UND GRUPPE  30 (N =  1), FEHLER =   0.0316
34 GRUPPEN NACH VEREINIGUNG VON GRUPPE  12 (N =  1) UND GRUPPE  28 (N =  1), FEHLER =   0.0664
33 GRUPPEN NACH VEREINIGUNG VON GRUPPE  25 (N =  1) UND GRUPPE  27 (N =  1), FEHLER =   0.1072
32 GRUPPEN NACH VEREINIGUNG VON GRUPPE  16 (N =  1) UND GRUPPE  35 (N =  1), FEHLER =   0.1361
31 GRUPPEN NACH VEREINIGUNG VON GRUPPE   3 (N =  1) UND GRUPPE  32 (N =  1), FEHLER =   0.1623
30 GRUPPEN NACH VEREINIGUNG VON GRUPPE  13 (N =  1) UND GRUPPE  22 (N =  1), FEHLER =   0.2386
29 GRUPPEN NACH VEREINIGUNG VON GRUPPE   9 (N =  1) UND GRUPPE  25 (N =  2), FEHLER =   0.2554
28 GRUPPEN NACH VEREINIGUNG VON GRUPPE  15 (N =  1) UND GRUPPE  16 (N =  2), FEHLER =   0.2673
27 GRUPPEN NACH VEREINIGUNG VON GRUPPE  18 (N =  1) UND GRUPPE  29 (N =  1), FEHLER =   0.3016
26 GRUPPEN NACH VEREINIGUNG VON GRUPPE   6 (N =  1) UND GRUPPE   8 (N =  1), FEHLER =   0.3980
25 GRUPPEN NACH VEREINIGUNG VON GRUPPE   9 (N =  3) UND GRUPPE  20 (N =  1), FEHLER =   0.4100
24 GRUPPEN NACH VEREINIGUNG VON GRUPPE  19 (N =  1) UND GRUPPE  34 (N =  1), FEHLER =   0.4130
23 GRUPPEN NACH VEREINIGUNG VON GRUPPE  21 (N =  1) UND GRUPPE  33 (N =  1), FEHLER =   0.5289
22 GRUPPEN NACH VEREINIGUNG VON GRUPPE   5 (N =  1) UND GRUPPE   7 (N =  1), FEHLER =   0.5740
21 GRUPPEN NACH VEREINIGUNG VON GRUPPE  10 (N =  1) UND GRUPPE  17 (N =  1), FEHLER =   0.6087
20 GRUPPEN NACH VEREINIGUNG VON GRUPPE  15 (N =  3) UND GRUPPE  23 (N =  1), FEHLER =   0.7922
19 GRUPPEN NACH VEREINIGUNG VON GRUPPE   1 (N =  2) UND GRUPPE  12 (N =  2), FEHLER =   0.8004
18 GRUPPEN NACH VEREINIGUNG VON GRUPPE   2 (N =  1) UND GRUPPE  24 (N =  1), FEHLER =   0.8842
17 GRUPPEN NACH VEREINIGUNG VON GRUPPE   3 (N =  2) UND GRUPPE  18 (N =  2), FEHLER =   1.1651
16 GRUPPEN NACH VEREINIGUNG VON GRUPPE  15 (N =  4) UND GRUPPE  31 (N =  1), FEHLER =   1.2092
15 GRUPPEN NACH VEREINIGUNG VON GRUPPE   9 (N =  4) UND GRUPPE  19 (N =  2), FEHLER =   1.3152
14 GRUPPEN NACH VEREINIGUNG VON GRUPPE  13 (N =  2) UND GRUPPE  21 (N =  2), FEHLER =   1.3183
13 GRUPPEN NACH VEREINIGUNG VON GRUPPE   1 (N =  4) UND GRUPPE  26 (N =  1), FEHLER =   1.4523
```

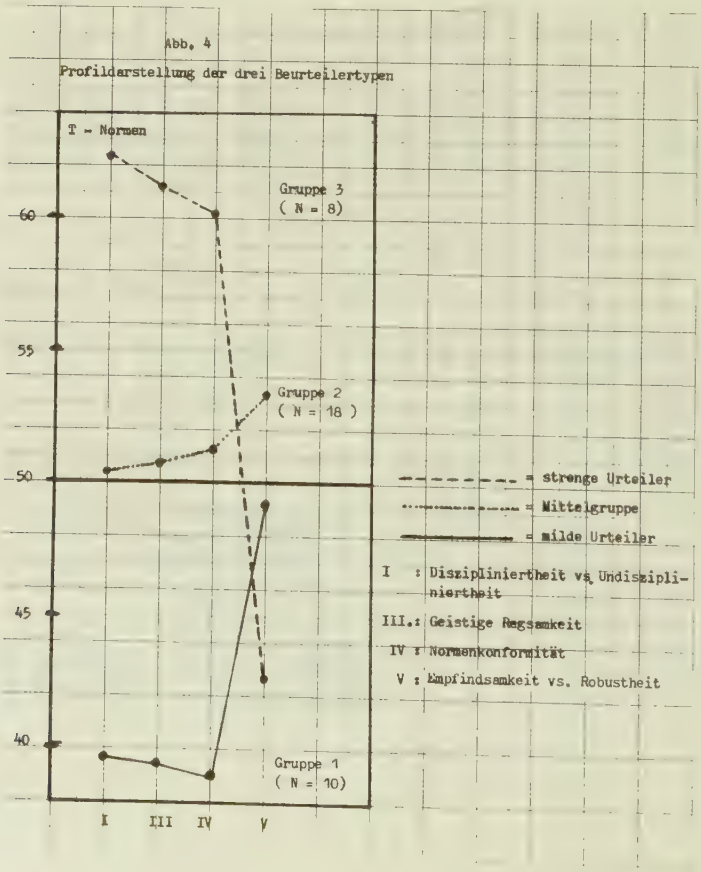
*Fortsetzung der Tab. 19*

```
 12 GRUPPEN NACH VEREINIGUNG VON GRUPPE   5 (N =  2) UND GRUPPE  36 (N =  1), FEHLER =    2.2136

 11 GRUPPEN NACH VEREINIGUNG VON GRUPPE   2 (N =  2) UND GRUPPE  10 (N =  2), FEHLER =    2.4660

 10 GRUPPEN NACH VEREINIGUNG VON GRUPPE   9 (N =  6) UND GRUPPE  14 (N =  1), FEHLER =    2.7148

  9 GRUPPEN NACH VEREINIGUNG VON GRUPPE   5 (N =  3) UND GRUPPE  13 (N =  4), FEHLER =    3.7978

  8 GRUPPEN NACH VEREINIGUNG VON GRUPPE   1 (N =  5) UND GRUPPE  11 (N =  1), FEHLER =    4.1760

  7 GRUPPEN NACH VEREINIGUNG VON GRUPPE   5 (N =  7) UND GRUPPE   9 (N =  7), FEHLER =    4.7860

  6 GRUPPEN NACH VEREINIGUNG VON GRUPPE   4 (N =  1) UND GRUPPE   6 (N =  2), FEHLER =    5.1646

  5 GRUPPEN NACH VEREINIGUNG VON GRUPPE   4 (N =  3) UND GRUPPE  15 (N =  5), FEHLER =    7.2323
GRUPPE  1 (N=  6) 01    11    29    31    33
GRUPPE  2 (N=  4) 02    10    17    27
GRUPPE  3 (N=  4) 03    19    32    35
GRUPPE  4 (N=  8) 04    06    08    15    16    26    34    39
GRUPPE  5 (N= 14) 05    07    09    13    14    20    21    22    23    28    30    36    38    40

  4 GRUPPEN NACH VEREINIGUNG VON GRUPPE   1 (N =  6) UND GRUPPE   2 (N =  4), FEHLER =   12.4406
GRUPPE  1 (N= 10) 01    02    10    11    12    17    27    29    31    33
GRUPPE  3 (N=  4) 03    19    32    35
GRUPPE  4 (N=  8) 04    06    08    15    16    26    34    39
GRUPPE  5 (N= 14) 05    07    09    13    14    20    21    22    23    28    30    36    38    40

  3 GRUPPEN NACH VEREINIGUNG VON GRUPPE   3 (N =  4) UND GRUPPE   5 (N = 14), FEHLER =   14.7270
GRUPPE  1 (N= 18) 01    02    10    11    12    17    27    29    31    33
GRUPPE  3 (N= 18) 03    05    07    09    13    14    19    20    21    22    23    28    30    32    35
                  36    38    40
GRUPPE  4 (N=  8) 04    06    08    15    16    26    34    39

  2 GRUPPEN NACH VEREINIGUNG VON GRUPPE   1 (N = 18) UND GRUPPE   3 (N = 18), FEHLER =   25.4169
GRUPPE  1 (N= 28) 01    02    03    05    07    09    10    11    12    13    14    17    19    20    21
                  22    23    27    28    29    30    31    32    33    35    36    38    40
GRUPPE  4 (N=  8) 04    06    08    15    16    26    34    39

STOP

1 MIN 17 SEC THIS SECTION  (0 MIN 34 SEC)
```

Die Darstellung des Ergebnisses der 2. Clusteranalyse soll zusätzlich graphisch erfolgen. Wir wählen eine Profildarstellung (Abbildung 4) der drei Beurteilertypen aufgrund der Gruppenmittelwerte bezüglich der vier Urteilsdimensionen Disziplin iertheit vs. Undisziplin iertheit (I), geistige Regsamkeit (III), Normenkonformität (IV) und Empfindsamkeit vs. Robustheit (V). Um die Gruppenmittelwerte zwischen den vier Urteilsdimensionen vergleichen zu können, wurden die T-Werte der Gruppenmittel aufgrund der T-Daten in Tabelle 17 errechnet.

Abb. 4

Profildarstellung der drei Beurteilertypen



Das Ergebnis der 2. Clusteranalyse kann als erfolgreich bezeichnet werden, denn alle Mittelwertsdifferenzen hinsichtlich der Dimension Disziplinierthe(en)it vs. Undisziplinierthe(en)it, geistige Regsamkeit und Normenkonformität lassen sich als signifikant ausweisen. Lediglich für die Dimension Empfindsamkeit vs. Robustheit ergibt sich nur eine signifikante Mittelwertsdifferenz, die zwischen Gruppe 2 und 3.

Fassen wir das Ergebnis der 2. Clusteranalyse interpretativ zusammen, so läßt sich feststellen: Gruppe 3 (N = 8) weist extrem hohe durchschnittliche Punktwerte auf den Dimensionen Disziplinierthe(en)it vs. Undisziplinierthe(en)it, geistige Regsamkeit und Normenkonformität auf, Gruppe 1 (N = 10) dagegen sehr niedrige. Bezüglich der drei genannten Urteilsdimensionen neigt Gruppe 3 also zu ungünstigen (strengen) und Gruppe 1 zu günstigen (milden) Schülerurteilen. Gruppe 2 (N = 18) gibt bezüglich der drei genannten Dimensionen Durchschnittsurteile ab. Die Dimension Empfindsamkeit vs. Robustheit dagegen erweist sich als weniger trennwirksam. Gruppe 2 neigt dazu, die Schüler als robuster und unkomplizierter zu charakterisieren als Gruppe 1 und insbesondere Gruppe 3. Als Hauptergebnis weist also die 2. Clusteranalyse zwei Extremgruppen und eine Mittelgruppe von Lehrerbeurteilern aus, wobei Gruppe 3 das implizite Persönlichkeitsmodell der Lehrer zur aktuellen Schülerbewertung im Sinne der Dimension Disziplinierthe(en)it vs. Undisziplinierthe(en)it, geistige Regsamkeit und Normenkonformität besonders streng und Gruppe 1 besonders milde realisiert.

Frage 11: Bei welchen Lehrerbeurteilertypen weichen die geschätzten (vermuteten) Schülerurteile von den tatsächlich erhaltenen Schülerurteilen in auffälligem Ausmaß ab?

Um diese Frage bearbeiten zu können, werden die auf die Fragebogenendform (vergl. Kapitel 7.3.5.5.) bezogenen Summenscores der Real- und Schätzurteile herausgegriffen und tabellarisch angeordnet. Bei der Gruppierung des Zahlenmaterials wird das Ergebnis der 2. Clusteranalyse berücksichtigt, demzufolge sich drei unterscheidbare Lehrerbeurteilertypen ergeben haben. Wir führen in Tabelle 20 die Summenscores der Lehrer bezüglich der Real- und Schätzurteile nach der erhaltenen Lehrerbeurteilertypologie zur aktuellen Schülerbewertung auf, d.h. also nach der Lehrergruppierung aufgrund der 2. Clusteranalyse.

Angesichts der Inspektion der Daten der Tabelle 20 fällt auf, daß Gruppe 1, die milde urteilenden Lehrer, ihre Schülerbewertung überschätzten und Gruppe 3, die streng urteilenden Lehrer, ihre Schülerbewertung unterschätzten (vergl. die Mittelwerte der Real- und Schätzurteile dieser beiden Gruppen).

Mit dem t - Test für Paardifferenzen (s. HOFSTÄTTER a.a.O.) prüfen wir diese Unterschiede auf Signifikanz. Die t-Tests wurden wieder auf der PDP 8 - I mit einem Programm von IHM gerechnet. Unter Zugrundelegung einer einseitigen Fragestellung erhalten wir für Gruppe 1 (die milden Urteiler) zwischen den Real- und Schätzwerten bei 9 Freiheitsgraden ein auf dem 5 % -Niveau signifikanten $t_{corr}$ - Wert von -2,09; für Gruppe 3 (die strengen Urteiler) bei 7 Freiheitsgraden ein auf dem 1 %-Niveau signifikanten $t_{corr}$ - Wert von +3,33. Milde Urteiler überschätzen also das Ausmaß angebrachten Interaktionsverhaltens aus der Sicht ihrer Schüler, und strenge Urteiler unterschätzten es.

Frage 12: Neigen günstig und ungünstig eingestufte Lehrer zu unterschiedlichen Schülerbeurteilungen?

Diese Frage läßt sich dahingehend spezifizieren, ob sich

Tab. 20

Real- und Schätzurteile bezüglich der Fragebogen-Endform für die drei Lehrerbeurteilertypen

| Lehrer Nr. | Realurteile | Schätzurteile | |
|---|---|---|---|
| 1 | 76.6 | 87 | |
| 2 | 79.6 | 85 | |
| 10 | 77.5 | 79 | |
| 11 | 78.3 | 76 | |
| 12 | 81.2 | 86 | Gruppe 1 ( milde Urteiler) |
| 17 | 81.7 | 89 | |
| 27 | 56.9 | 78 | |
| 29 | 52.6 | 63 | |
| 31 | 79.2 | 69 | |
| 33 | 76.7 | 83 | |
| $\bar{X}$ | 74.0 | 79.5 | |
| 3 | 64.7 | 58 | |
| 5 | 76.5 | 66 | |
| 7 | 82.3 | 67 | |
| 9 | 77.6 | 74 | |
| 13 | 72.6 | 77 | |
| 14 | 75.7 | 78 | |
| 19 | 70.6 | 74 | |
| 20 | 76.9 | 63 | Gruppe 2 ( Mittelgruppe) |
| 21 | 71.3 | 80 | |
| 22 | 72.4 | 71 | |
| 23 | 84.9 | 79 | |
| 28 | 86.9 | 78 | |
| 30 | 83.5 | 88 | |
| 32 | 72.3 | 80 | |
| 35 | 72.0 | 72 | |
| 36 | 50.7 | 63 | |
| 38 | 78.8 | 78 | |
| 40 | 67.0 | 72 | |
| $\bar{X}$ | 74.2 | 73.2 | |
| 4 | 79.6 | 63 | |
| 6 | 78.1 | 80 | |
| 8 | 67.7 | 59 | |
| 15 | 83.7 | 76 | Gruppe 3 ( strenge Urteiler) |
| 16 | 84.0 | 81 | |
| 26 | 81.2 | 79 | |
| 34 | 81.8 | 75 | |
| 39 | 84.9 | 75 | |
| $\bar{X}$ | 80.1 | 73.5 | |

aufgrund der Einstufung durch die Schüler über dem Median liegende von unter dem Median liegenden Lehrern bezüglich der fünf Beurteilungsdimensionen unterscheiden bzw. aufgrund der fünf Beurteilungsdimensionen diskriminieren lassen.

Die 36 Lehrer, die Schülerbeurteilungen anhand der Polaritäten abgegeben haben, werden nach ihren Einstufungen aus der Sicht der Schüler wieder in zwei Hälften geteilt, in eine über und eine unter dem Median liegende Gruppe. Tabelle 21 gibt die durchschnittlichen Beurteilungsscores dieser beiden Gruppen auf den fünf Beurteilungsdimensionen zur aktuellen Schülerbeurteilung wieder. Für diejenigen Lehrer, die von ihren Schülern über dem Median liegend eingestuft wurden, weist Tabelle 21 höhere Mittelwerte bezüglich der Dimensionen I, II, III und IV aus.

Die obige Hauptuntersuchungsfrage, ob die fünf Urteilsdimensionen zwischen über und unter dem Median liegenden Lehrern diskriminieren, läßt unter anderem mit Hilfe der Diskriminanzanalyse (DA) eine Antwort zu (COOLEY UND LOHNES, 1962, S. 116 f; WEBER, 1967, S. 523 f).

Der Zweck der DA ist die Bestimmung der Zugehörigkeit eines Merkmalsträgers zu einer von mehreren Kriteriumsklassen aufgrund mehrerer Prädiktorvariablen. Die Vpn werden zunächst auf diskrete Klassen aufgeteilt, in unserem Fall auf die entsprechenden Gruppen oberhalb und unterhalb des Medians. Als Prädiktorvariablen haben wir die fünf Urteilsdimensionen zur aktuellen Schülerbeurteilung.

Bei Anwendung der einfachen linearen DA nach R.A. FISHER (vergl. WEBER, 1967, a.a.O.) wird nur eine Diskriminanzfunktion gesucht, mit der sich der Diskriminanzscore

Tab. 21

Durchschnittliche Beurteilungsscores der aufgrund der Einstufung durch die Schüler über und unter dem Median liegenden Lehrer bezüglich der fünf Urteilsdimensionen zur aktuellen Schülerbewertung

| Lehrer Nr. | I | II | III | IV | V | Diskriminanzanalytische Rechenmaße* | |
|---|---|---|---|---|---|---|---|
| 2 | 21.2 | 21.3 | 10.2 | 7.2 | 8.7 | 9.9 | |
| 4 | 34.4 | 28.2 | 18.6 | 15.2 | 4.6 | 9.9 | |
| 6 | 33.1 | 28.3 | 16.9 | 10.9 | 6.2 | 8.0 | |
| 9 | 27.2 | 30.2 | 15.0 | 11.1 | 8.6 | 5.7 | |
| 11 | 21.9 | 26.3 | 9.3 | 9.3 | 4.1 | 8.5 | Gruppe der in der Schülerbewertung über dem Median liegenden Lehrer |
| 12 | 23.7 | 28.5 | 13.0 | 10.1 | 7.8 | 10.1 | |
| 15 | 28.8 | 28.4 | 17.4 | 13.4 | 7.8 | 10.8 | |
| 16 | 28.9 | 29.3 | 15.8 | 13.4 | 7.5 | 11.3 | |
| 17 | 21.5 | 26.4 | 10.4 | 9.9 | 9.6 | 10.1 | |
| 23 | 28.9 | 25.7 | 14.1 | 10.8 | 7.0 | 10.6 | |
| 26 | 32.5 | 28.3 | 17.3 | 13.5 | 7.8 | 12.3 | |
| 28 | 28.1 | 26.7 | 14.9 | 10.2 | 8.9 | 10.7 | |
| 30 | 27.7 | 29.7 | 15.2 | 10.5 | 9.4 | 10.8 | |
| 31 | 24.1 | 27.7 | 12.4 | 9.9 | 7.5 | 8.5 | |
| 34 | 29.3 | 28.0 | 14.6 | 14.7 | 7.3 | 13.6 | |
| 38 | 29.9 | 24.3 | 14.8 | 11.6 | 8.9 | 6.4 | |
| 39 | 29.2 | 30.4 | 16.2 | 13.3 | 8.1 | 14.6 | |
| 7 | 26.7 | 27.8 | 16.3 | 11.7 | 7.3 | 10.2 | |
| $\bar{X}$ | 27.6 | 27.5 | 14.6 | 11.5 | 7.2 | 10.1 | |
| 1 | 25.5 | 21.0 | 12.3 | 10.1 | 6.5 | 12.5 | |
| 3 | 24.6 | 28.4 | 11.8 | 14.3 | 7.3 | 9.2 | |
| 5 | 26.6 | 23.5 | 17.9 | 10.8 | 8.1 | 9.0 | |
| 8 | 32.6 | 31.2 | 15.1 | 11.5 | 6.4 | 5.6 | |
| 10 | 18.7 | 25.9 | 11.9 | 9.2 | 9.9 | 11.1 | |
| 13 | 27.7 | 27.3 | 14.3 | 10.1 | 7.6 | 9.0 | |
| 14 | 25.8 | 24.5 | 13.7 | 10.4 | 10.7 | 8.9 | |
| 19 | 22.9 | 24.3 | 11.4 | 13.0 | 8.9 | 8.3 | |
| 20 | 13.0 | 26.1 | 14.2 | 11.3 | 7.9 | 8.0 | |
| 21 | 27.4 | 29.7 | 13.4 | 10.6 | 8.9 | 7.9 | Gruppe der in der Schülerbewertung unter dem Median liegenden Lehrer |
| 22 | 25.2 | 30.7 | 13.1 | 11.3 | 7.6 | 7.5 | |
| 27 | 24.2 | 24.2 | 12.5 | 7.6 | 8.7 | 10.8 | |
| 29 | 20.7 | 29.0 | 12.0 | 9.1 | 6.4 | 10.3 | |
| 32 | 23.1 | 28.8 | 12.9 | 13.4 | 8.5 | 7.2 | |
| 33 | 25.4 | 28.7 | 12.2 | 10.2 | 6.8 | 9.5 | |
| 35 | 23.5 | 24.7 | 11.6 | 13.8 | 7.8 | 6.8 | |
| 36 | 28.5 | 24.6 | 13.5 | 12.7 | 7.9 | 9.4 | |
| 40 | 23.1 | 32.0 | 15.4 | 8.9 | 8.4 | 6.2 | |
| $\bar{X}$ | 25.4 | 26.9 | 13.3 | 11.0 | 8.0 | 8.7 | |

* Die diskriminanzanalytischen Rechenmaße wurden ermittelt aufgrund der Trennfunktion $Y_k = 0.30\ X_{1k} + 0.04\ X_{2k} + 0.46\ X_{3k} + 0.21\ X_{4k} - 0.45\ X_{5k}$

für jeden Meßwertträger aus jeder Kriteriumsklasse ermitteln läßt. Die Diskriminanzfunktion gibt die Klassenzugehörigkeit der Merkmalsträger in der Regel besser an als jede Einzelvariable. Der Diskriminanzscore der Vp K ergibt sich aus der Summe der (gewichteten) Rohwerte in den Prädiktorvariablen:

$$Y_k = b_1 X_{1k} + b_2 X_{2K} \ldots\ldots + b_n X_{nK}$$

In dieser Gleichung bedeuten:

$Y_k$ : der Diskriminanzscore der $Vp_k$
$b_j$ : die Gewichte der Prädiktorvariablen
$X_{jk}$ : die Ausprägung der Prädiktorvariablen j bei der $Vp_k$.

Unbekannt in dieser Gleichung sind die Gewichtskoeffizienten $b_j$. Sie werden so geschätzt, daß die Varianz der Trennwerte $Y_k$ zwischen den Kriteriumsklassen möglichst groß und innerhalb der Klassen möglichst klein ist.

Im Falle der multiplen DA ergeben sich mehrere orthogonale Diskriminanzfunktionen, die zur Trennung der Gruppen beitragen können.

Zur Beurteilung, ob die Durchführung einer DA sinnvoll gewesen ist, gehört die Vorbeurteilung: Die Durchführung einer DA hat dann Erfolg gehabt, wenn die Mittelwertsunterschiede zwischen den Kriteriumsklassen bezüglich der Trennvariablen (signifikant) größer sind als die Mittelwertsdifferenzen zwischen den Kriteriumsklassen bezüglich der "trennwirksamsten" Einzelvariablen.

Wir stellen die Nullhypothese auf, daß sich die beiden Lehrergruppen aufgrund der fünf Urteilsdimensionen diskriminanzanalytisch nicht voneinander trennen lassen. Wir rechneten eine einfache DA und erhielten folgende Trennfunktion:

$$Y_K = 0{,}30 X_{1K} + 0{,}04 X_{2K} + 0{,}46 X_{3K} + 0{,}21 X_{4K} - 0{,}45 X_{5K}.$$

Die DA setzt Normalität der einzelnen Prädiktoren in den Kriteriumsklassen und Varianzhomogenität der einzelnen Prädiktoren zwischen den Kriteriumsklassen sowie Unabhängigkeit der Kriteriumsklassen voneinander voraus. Diese Anforderungen an die Daten wurden überprüft und sind erfüllt.

Der durchschnittliche Trennwert der über dem Median liegenden Lehrer beträgt $\bar{X}^{A} = 10,1$, derjenige der unter dem Median liegenden Lehrer $\bar{X}^{B} = 8,7$ (vergl. Tabelle 21). Die Differenz der beiden Mittelwerte ist signifikant ($t_2 = 2,06$; $df = 34$; $p \leqslant 0,05$). Demgegenüber gestattet die Dimension I (Diszipliniertheit vs. Undiszipliniertheit), die den größten Mittelwertsunterschied aufweist, noch keine signifikante Trennung der beiden Gruppen voneinander ($t_1 = 1,86$; $df = 34$). Daß das Ergebnis der DA als erfolgreich betrachtet werden kann, wird durch $t_2$ ausgewiesen, das mit 2,06 einen höheren (signifikanten) kritischen Wert ergibt als $t_1$ mit 1,86.

Über dem Median liegende, also aus der Sicht der Schüler eher günstig eingestufte Lehrer, neigen zu strengeren Schülerurteilen im Sinne der Dimensionen zum impliziten Persönlichkeitsmodell als eher ungünstig eingestufte Lehrer.

## 9. Zusammenfassung der Ergebnisse der Hauptuntersuchung

1) Die dreifaktorielle Aufgliederung der Items aufgrund der Voruntersuchung wurde, den Ergebnissen einer zweiten R - Analyse zufolge, der eine Haupt-, Real-und Sonderschülerstichprobe von N = 937 zugrunde lag, auf zwei Faktoren Strenge und Unterstützung reduziert. Bezogen auf die Gesamtstichprobe (N = 937) ergaben sich für Strenge

und Unterstützung Reliabilitätskoeffizienten ($r_{stab}$) von 0,90 bzw. 0,85, bezogen auf die Sonderschülerstichprobe Koeffizienten von 0,80 bzw. 0,77.

2) Die an der Hauptuntersuchung beteiligten Lehrer schätzen das Ausmaß ihrer von den Schülern perzipierten Strenge besser ein als das Ausmaß an Unterstützung. Insgesamt kovariieren die Beurteilungsmodi (Real-, Schätz- und Wunschurteile) für Strenge höher als für Unterstützung. Besonders hohe Korrelationen ergeben sich zwischen den Schätz- und Wunschurteilen für Strenge, aber auch für Unterstützung.

3) Eine genaue Inspektion der faktoriellen Itemaufgliederung erbrachte, daß das Auffinden der beiden Faktoren Strenge und Unterstützung hauptsächlich durch den Umstand zweier verschiedener Schwierigkeitsclusters der Items ($\bar{p}_{Strenge}$ = 0,37; $\bar{p}_{Unterstützung}$ = 0,73) bedingt ist. Deshalb erschien es uns problematisch, weiterhin von homogenen Itemgruppen auf den unabhängigen Dimensionen Strenge und Unterstützung zu sprechen.

4) Zur Auswahl für die Endform des Schülerbeobachtungsbogens, die 25 Items umfaßt, wurden folgende Items berücksichtigt:

- Items, die keine oder nur geringe Korrelationen zu sozial erwünschtem Antwortverhalten aufweisen;
- Items, die sich, einem Expertenrating zufolge, in Bezug auf Klassenklima und Lernatmosphäre als trennwirksam charakterisieren lassen, die also in extrem positivem und extrem negativem Ausmaß zum Klassenklima beitragen;
- Items, für die zusätzlich hohe Trennschärfen, bezogen auf die Subskalen Strenge und Unterstützung, errechnet wurden;
- Items, die zusätzlich hohe F-Prüfgrößen aufweisen als Kriterium, daß sie zwischen den Lehrern differenzieren.

5) Für die Fragebogenendform zur Erfassung des Interaktions- und Unterrichtsstils von Lehrern aus der Sicht der Schüler, die 25 Items positiver und negativer Bekräftigungsweisen von Lehrern umfaßt (s. Tabelle 13 b), errechneten wir eine Reliabilität (nach KUDER - RICHARDSON) von 0,90 bezogen auf die Gesamtstichprobe und von 0,84 bezogen auf die Sonderschülerstichprobe.

Der durchschnittliche Retest-Koeffizient aus drei Lernbehindertenklassen liegt bei 0,85, der Retest-Koeffizient für eine Hauptschulklasse bei 0,96.

6) Das mit unserer Fragebogenendform erfaßte Ausmaß angebrachten Interaktionsverhaltens ist - soweit es unsere Stichprobe betrifft - bei Lehrern und Lehrerinnen in Sonderschulen für Lernbehinderte nicht von dem der Haupt- und Realschullehrer(innen) verschieden.

7) Lehrer, die von ihren Schülern eher günstig eingestuft werden, unterschätzen ihr Ausmaß angebrachten Interaktionsverhaltens aus der Sicht ihrer Schüler und Lehrer, die von ihren Schülern eher ungünstig eingestuft werden, überschätzen es.

8) Günstig beurteilte Lehrer weisen geringere Diskrepanzen zwischen ihren Real- und Wunschurteilen auf als ungünstig beurteilte Lehrer.

9) Im Sinne der den Lehrern eigenen impliziten Persönlichkeitstheorie zur aktuellen Schülerbeurteilung ließen sich mit Hilfe der Clusteranalyse drei klar unterscheidbare Lehrerbeurteilertypen bzw. -gruppen identifizieren, nämlich zwei Extremgruppen und eine Mittelgruppe. Bezüglich der Urteilsdimensionen Diszipliniertheit vs. Undiszipliniertheit, geistige Regsamkeit und Normenkonformität

realisiert die eine Extremgruppe besonders milde, die andere besonders strenge Schülerurteile. Für streng urteilende Lehrer gilt, daß sie ihr Ausmaß angebrachten Interaktionsverhaltens aus der Sicht ihrer Schüler unterschätzen, für milde urteilende Lehrer dagegen gilt, daß sie es überschätzen.

10) Aus der Sicht ihrer Schüler eher günstig eingestufte Lehrer neigen zu strengeren Schülerurteilen im Sinne der Dimensionen zum impliziten Persönlichkeitsmodell als eher ungünstig eingestufte Lehrer.

## 10. Diskussion der Ergebnisse der Hauptuntersuchung

Aus der Vielfalt der Wirkungszusammenhänge, die sich in Schulklassen zwischen Lehrern und Schülern abspielen, haben wir, einigen lernpsychologischen Überlegungen folgend, einen relativ kleinen, aber fest umrissenen Bereich, nämlich das Belohnungs- und Bestrafungsverhalten von Lehrern aus der Sicht der Schüler, herausgegriffen und analysiert.

Die Bestimmung der Dimensionalität eines großen Itempools erfolgte zunächst faktorenanalytisch in Anlehnung an ein zweidimensionales Modell elterlicher Bekräftigungsweisen Strenge und Unterstützung nach HERRMANN et al. (1971). Unter Strenge und Unterstützung wurde - bezogen auf die konkrete Erziehungssituation - die Förderung von erwünschten bzw. die Unterbindung von unerwünschten Schüler - Verhaltensweisen mit verbalen und instrumentellen Mitteln verstanden. Strenge und Unterstützung wurden operational mit Fragebogen - Skalen erfaßt, die die Bekräftigungswirkung von Lehrern aus der Sicht der Schüler auf einer molaren Verhaltensebene charakterisieren sollten.

Vor- und Hauptuntersuchung stellten heraus, daß die Lehrer das Ausmaß ihrer von den Schülern perzipierten Strenge besser einschätzen konnen als ihr Ausmaß an Unterstützung. Insgesamt kovariieren die Beurteilungsmodi (Real-, Schätz- und Wunschurteile) für Strenge höher als für Unterstützung. Besonders hohe Korrelationen erhielten wir zwischen den Schätz- und Wunschurteilen für Strenge und Unterstützung. Demgegenüber fallen die Korrelationen zwischen den Real- und Schätzurteilen einerseits und zwischen den Real- und Wunschurteilen andererseits für die Unterstützungsdimension deutlich ab. Dieser letzte Befund gilt jedoch nicht für die Strengedimension. Die unterschiedlich hohen Korrelationen zwischen den Beurteilungsmodi für Strenge und Unterstützung interpretieren wir dahingehend, daß die Lehrer kaum imstande sind, im Unterstützungsbereich die Schätz- und Wunschurteile an den Realurteilen auszurichten. Hieraus ergibt sich die Konsequenz, beim Verhaltenstraining von Lehrern eine ständige Rückmeldung und Unterweisung im Hinblick auf die Handhabung positiver sozialer Verstärker vorzunehmen.

Wir hatten die Faktorenanalyse zunächst zur Bestimmung der Dimensionalität des Itempools herangezogen. Die Faktorenanalyse erfüllte dabei eine vorläufige Ordnungsfunktion, der gesamte Itempool wurde in zwei homogene Itemgruppen aufgegliedert, die wir in Anlehnung an die Marburger Erziehungsskalen zur Erfassung des elterlichen Erziehungsstils (vergl. HERRMANN et al., 1971) mit Strenge und Unterstützung bezeichneten.

Bei HERRMANN et al. (1971) werden die beiden Modellkomponenten Strenge und Unterstützung als unabhängig voneinander postuliert. Hierbei handelt es sich nach STAPF et al. (1971) "zunächst um eine t h e o r e t i s c h g e s e t z t e interne Modellrelation zwischen den beiden Modellkomponenten Strenge und Unterstützung.

Davon zu unterscheiden ist das methodische Problem der empirischen Überprüfung dieser Unabhängigkeitsbeziehung." In unserem Falle stellte sich angesichts der zweifaktoriellen Itemaufgliederung heraus, daß das Auffinden der beiden Faktoren Strenge und Unterstützung auf den Umstand zweier verschiedener Schwierigkeitsclusters der Items, also auf die unterschiedliche Beantwortungshäufigkeit von Strenge- und Unterstützungsitems durch die Schülerstichprobe zurückzuführen ist. Strenge- und Unterstützungsitems mit homogenen (d.h. nicht unterschiedlichen) Schwierigkeiten würden sich vermutlich auf einer gemeinsamen (bipolaren) Dimension lokalisieren lassen, wodurch wir eine gemischte Skala aus Strenge- und Unterstützungsitems erhalten hätten. Deshalb erschien es uns problematisch, weiterhin von homogenen Itemgruppen auf den unabhängigen Dimensionen Strenge und Unterstützung zu sprechen.

Ein Expertenrating erbrachte als externes Selektionsverfahren Aufschlüsse darüber, welche Items des oben genannten Itempools sich am besten zur Charakterisierung angebrachten Interaktionsverhaltens von Lehrern, bezogen auf Lernatmosphäre und Klassenklima, eignen. Dadurch reduzierten wir das zuvor kritisierte zweidimensionale Konzept Strenge und Unterstützung auf e i n e Dimension, die wir in Anlehnung an unsere lernpsychologische Terminologie "Bekräftigungswirkung des Lehrers" nennen und die sowohl Lohn- als auch Strafitems enthält.

Für die Endform des Schülerfragebogens wurden aus dem oben genannten Itempool solche Items berücksichtigt, die

a) dem Ergebnis des Expertenratings zufolge in extrem positivem und extrem negativem Ausmaß zum Klassenklima beitragen;
b) die Tendenz zu sozial erwünschtem Antwortverhalten weitgehend ausschließen;
c) in Bezug auf perzipiertes Belohnungs- und Bestrafungsverhalten von Lehrern hohe Trennschärfen aufweisen;

d) Lehrer voneinander differenzieren und
e) für Real-, Haupt- und lernbehinderte Sonderschüler siebter, achter und neunter Schuljahre eindeutig beobachtbar und registrierbar sind.

Für die Fragebogenendform zur Erfassung des Interaktions- und Unterrichtsstils von Lehrern aus der Sicht der Schüler, die 25 Items positiver und negativer Bekräftigungsweisen von Lehrern umfaßt (vergl. Tabelle 13 b), errechneten wir eine Reliabilität (innere Konsistenz nach KUDER-RICHARDSON 21) von 0,90 bezogen auf die Gesamtstichprobe und von 0,84 bezogen auf die Sonderschülerstichprobe. Diese beiden hohen internen Konsistenz-Koeffizienten und die hohen Trennschärfen der Items der Fragebogenendform legen nahe, daß Lohn- und Strafitems der Fragebogenendform dasselbe messen und lassen es gerechtfertigt erscheinen, Lohn- und Strafitems zu e i n e r Fragebogenendform, d.h. zu e i n e r Dimension zusammenzufassen.

Die hohen Retest - Koeffizienten der Fragebogenendform von 0,85, bezogen auf drei Lernbehindertenklassen, und 0,96, bezogen auf eine Hauptschulklasse, weisen darauf hin, daß relativ stabile Verhaltensmerkmale über die Zeit erfaßt wurden. Hier zeigt sich der Vorteil der Methode der Verhaltenseinschätzung von Lehrern durch Schüler, nämlich daß die Reliabilität der Schätzurteile von Schülern über Lehrer im Sinne der SPEARMAN - BROWN - Formel eine Funktion der Anzahl der Beurteiler ist. Da sich die Itemmittelwerte bei ca. 30 und mehr Schülern pro Klasse stabilisieren, werden die Schätzurteile so reliabel wie Schulleistungs- und Intelligenztests. Das leichte Absinken der Reliabilitätskoeffizienten für die Sonderschülerstichprobe läßt sich vermutlich auf den Umstand zurückführen, daß in Lernbehindertenklassen weit weniger Schüler sitzen als in Haupt- und Realschulklassen. Dadurch ist die Anzahl der Beurteiler pro Lehrer in Lernbehindertenklassen stark reduziert.

Bei Inspektion der Rohpunkteverteilung der Fragebogenendform (vergl. Abbildung 3) fällt auf, daß sich eine rechtsgipflige Verteilung andeutet. Diese Rechtsgipfligkeit kann im Falle unserer Untersuchung durch einen Selektionseffekt bedingt sein. Es liegt nahe, anzunehmen, daß bei Lehrereinschätzungen durch Schüler bevorzugt erfolgreiche Lehrer teilnehmen, die sich ihren erfolgreichen Unterrichtsstil gern von ihren Schülern bescheinigen lassen. Der Vorteil einer rechtsgipfligen Verteilung kann andererseits darin erblickt werden, daß die Lehrer im unteren Bereich der Verteilung (also von ihren Schülern eher ungünstig eingestufte Lehrer) besser differenziert werden.

In der Voruntersuchung, die zwölf Haupt- und vierzehn Gesamtschullehrer einbezog, vermuteten die Lehrer, von ihren Schülern strenger und weniger unterstützend eingestuft zu werden. Kennzeichnet man ein hohes Ausmaß an Unterstützung und ein geringes Ausmaß an Strenge als angebrachtes Interaktionsverhalten, so neigen die in der Voruntersuchung von ihren Schülern beurteilten Lehrer dazu, "ihr Licht unter den Scheffel zu stellen."

Dieser Sachverhalt zeigt sich in der Hauptuntersuchung, die eine weitaus größere Lehrerstichprobe umfaßte, jedoch nur für die von ihren Schüler als günstig eingestufte Lehrergruppe, während die unterhalb des Median liegende Lehrergruppe ihr Ausmaß angebrachten Interaktionsverhaltens im Urteil ihrer Schüler überschätzt. Die Schätzurteile der oberhalb und unterhalb des Median befindlichen Lehrergruppen weisen gewissermaßen einen Trend zur Mitte auf bzw. richten sich zur "zentralen Tendenz" aus. Die ungünstig beurteilte Lehrergruppe (unterhalb des Median liegend), die weniger Lob, Hilfe und Förderung und mehr Bestrafungsstrategien praktiziert, unterliegt bei der Einschätzung ihres Belohnungs- und Bestrafungsverhaltens in der Sicht ihrer Schüler möglicherweise zahlreichen unmittelbaren

subjektiven Effektivitätserfahrungen, die dadurch bedingt sind, daß überwiegend strafendes Verhalten von Lehrern in Konfliktsituationen signifikant mit s o f o r t i g e r Anpassung der Schüler korreliert, aber langfristig mit einer Verschlechterung des sozial-emotionalen Klimas in der Klasse einhergeht (vergl. TAUSCH, 1970, S. 233). Bei ungünstig eingestuften Lehrern könnten vorhandene Widerstände gegenüber Verhaltensänderungen, Widerstände, die durch die genannten unmittelbaren subjektiven Effektivitätserfahrungen infolge überwiegend bestrafender Strategien bedingt sind, dadurch abgebaut werden, indem man solchen Lehrern ihre überaus große Diskrepanz zwischen ihren Real- und Wunschurteilen vor Augen führt.

Hinsichtlich der den Lehrern eigenen impliziten Persönlichkeitstheorie zur aktuellen Schülerbeurteilung erscheinen zwei extreme Lehrergruppen mit unterschiedlichen Schülerbeurteilungstendenzen von Interesse. Im Sinne der Urteilsdimensionen Diszipliniertheit vs. Undiszipliniertheit, geistige Regsamkeit und Normenkonformität realisiert die eine Extremgruppe besonders milde, die andere Extremgruppe besonders strenge Schülerurteile. Offenbar lassen sich diese beiden Lehrergruppen bei der Beurteilung der Schülerpersönlichkeit von zwei Konzepten leiten, die sich angesichts sozialpsychologischer bzw. gruppendynamischer Befunde nicht zur Deckung bringen lassen und mit der Divergenz bei Führerrollen zusammenhängen, der Divergenz bei aufgabenbezogenen und sozial-emotional orientierten Gruppenleitern (s. HOFSTÄTTER, 1964, S. 130 f.). Demnach lassen sich die strengen Beurteiler der Schülerpersönlichkeit eher von einem aufgaben- und stoffplanorientierten Effektivitätskonzept und die milden Beurteiler eher von einem Zufriedenheitskonzept für ihre Schüler leiten. Dieser Sachverhalt mag dazu beitragen, daß strenge Urteiler das Ausmaß angebrachten Interaktionsverhaltens in der Sicht ihrer Schüler unter- und milde Urteiler es überschätzen.

Literaturverzeichnis

ASCHERSLEBEN, K.: Entwicklung eines Lügen-Scores zur Messung von Simulationstendenzen. Zt. f. Entwicklungspsychol. und Pädagogische Psychol. 1972, Bd. IV, Heft 2, 118 - 129.

ATKINSON, Rita & Nancy M. ROBINSON: Paired-associate learning by schizophrenic and normal subjects under conditions of personal and inpersonal reward and punishment. J. abnorm. soc. Psychol. 1961, 62, 321 - 326.

AUSUBEL, D.P. et al.: Perceived Attitudes as determiners of children's ego structure. Cild. Develp. 1954, 25, 173 - 184.

AZRIN, N.H. & W.C. HOLZ: Punisment. In Honig, W.K. (Ed.): Operant behavior: Areas of research and Application. New York: Appleton - Century - Crofts 1966, 380 - 447.

BARTRAM, M.: Konstruktion eines Fragebogens zur Erfassung des Interaktionsstils von Lehrern. In Vorbereitung, 1972.

BARTRAM, M., ROLLETT, B. und SUHRE: Kriterien der Schülerbeurteilung von Lehrern. Unveröffentlichte Untersuchung an der PH Osnabrück, 1971.

BECKER, W.C.: The parent attitudes research instrument - a research review. Child Develp. 1965, 36, 329 - 365.

BLÖSCHL, Lilian: Belohnung und Bestrafung im Lernexperiment. Weinheim, Basel 1969.

BORGATTA, E.F. & D.C. GLASS: Personality concomitants of extreme response set. J. soc. Psychol. 1961, 55, 213 - 221.

BROWN, A.W., MORRISON, J. & G.B. COUCH: Influence of effectional family relationship on character developments. J. abnorm. soc. Psychol. 1947, 42, 422 - 428.

BRUNER, J.S. & TAGUIRI, R.: The perception of people, in Lindzey, G. (Ed.): Handbook of Social Psychology, Vol. 2 Cambridge, Mass., 1954, 634 - 654.

BRUNING, J.L. & B.L. KINTZ: Computational Handbook of Statistics Glenview, Illinois 1968, 54 - 60.

CATTELL, R.B.: The scientific analysis of personality. Hammondsworth/Middlesex: Penguin Books, 1965.

COOLEY, W. & P.R. LOHNES: Multivariate procedures for the Behavioral Sciences. New. York: Wiley 1962.

CRONBACH, L.J.: Studies of acquiescence as a function in the measurement of personality. Psychol. Bull. 1938, 35, 659.

CRONBACH, L.J.: Response sets and test validity. Educ. Psychol. Measurement 1964, 6, 475 - 494.

CRONBACH, L.J.: Processes affecting scores on "understanding of others" and "assumed similarity". Psychol. Bull. 1955, 52, 177 - 193.

CRONBACH, L.J.: Essentials of psychological testing. New York, London: Harper & Row 1964.

CURRY, C.R.: Case Studies in Behavior Modification. Psychol. in the School, 7, 1970, 330 - 335.

DAMARIN, F. & MESSICK: Response Styles as Personality Variables: A theoretical Integration of Multivariate Research, Princeton 1965.

DAHLSTROM, W.G.: The roles of social desirability and acquiescence in responses to the MMP I. In: MESSICK, S. & J. ROSS (Eds): Measurement in personality and cognition. New York 1962.

EDWARDS, A.L.: The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. J. apl. Psychol. 1953, 37, 90 - 93.

EDWARDS, A.L.: The social desirability variable in personality assessment and research. New York 1957.

ESTES, W.K.: An experimental study of punishment. Psychol. Monogr. 1944, 57, No. 263.

ESTES, W.K.: Outline of a theory of punishment. In: CAMPBELL, B.A. & R.M. CHURCH (Eds): Punishment and aversiv behavior. New York: Appleton - Century - Crofts 1969, 57 - 82.

EYFERTH, K.: Methoden zur Erfassung von Erziehungsstilen. In: HERRMANN, T. (Hrsg): Psychologie der Erziehungsstile. Göttingen 1966, 17 - 31.

EYSENCK, H.J.: Maudsley - Persönlichkeitsfragebogen. Göttingen 1964

FISCHER, G.H.: Neue Entwicklungen in der psychologischen Testtheorie. In: FISCHER, G.H. (Hrsg): Psychologische Testtheorie. Bern 1968, 13 - 158.

FLANDERS, N.A.: Interaction Analysis in the classroom. A Manual for Observers. College of Education, University of Minnesota, Minneapolis 1960.

GAGE, N.L., RUNKEL. P.J. & B.B. CHATTERJEE: Equilibrium theory and behavior change: An experiment in feedback from pupils to teachers. Urbana: Bur. of Educ. Res., Univers. of Illinois 1960.

GRUBITSCH, S. und H. VOGT: Das Polaritätsprofil zur Erfassung stereotyper Urteilstrukturen aus der Sicht 11/12- und 15/16 jähriger Volksschüler. Schule und Psychol. 1970, 4, 97 - 112; 5, 144 - 150.

GRUBITSCH, S.: Entwurf eines Beobachtungsverfahrens zur Erfassung der Lernsituation und Lernatmosphäre im Unterricht (BELU - V - 1 - D). Unveröffentlicht, Oldenburg 1970.

GUILFORD, J.P.: Persönlichkeit. Weinheim 1965$^2$.

GUTTMAN, L.A.: The basis of scalogramm analysis. in: STOUFFER, S.A. et al. (Eds): Measurement and prediction. New York: Wiley 1950, 60 - 90.

GUTTMAN, L.A.: A generalized simplex for factor analysis. Psychometrika 1955, 20, 173 - 192.

HAYS, W.L.: An approach to the study of trait implication and trait similarity. In: TAGUIRI, R. & L. PETRULLO (Eds): Person perception and interpersonal behavior. Stanford: Univ. Press. 1958, 289 - 299.

HERRMANN, T.: Urteilsnuanciertheit und Fraktionierung des Bezugssystems. Psychol. Beitr. 1964, 7, 539 - 557.

HERRMANN, T., SCHWITAJEWSKI, E. & H.J. AHRENS: Untersuchungen zum elterlichen Erziehungsstil: Strenge und Unterstützung. Arch. ges. Psychol. 1968, 120, 74 - 105.

HERRMANN, T.: Lehrbuch der empirischen Persönlichkeitsforschung. Göttingen 1969.

HERRMANN, T., STAPF, A. und W. KROHNE: Die Marburger Skalen zur Erfassung des elterlichen Erziehungsstils. Diagnostica 1971, 17, 118 - 133.

HOFER, M.: Die Schülerpersönlichkeit im Urteil des Lehrers. Weinheim, Basel, Berlin 1969.

HOFSTÄTTER, P.R.: Gruppendynamik. Rowohlt, München 1964.

HOFSTÄTTER, P.R. und D. WENDT: Quantitative Methoden der Psychologie. München 1966.

HORST, P.: Psychological measurement and prediction. Belmont, Calif.: Wadsworth Publ. Comp. 1966.

HUGHES, M.: Development of the means for the assessment of the quality of teaching in elementary schools. Salt Lake City: Univ. of Utah Press 1959.

JAYNE, C.D.: A study of the relationship between teaching and educational outcomers. J. exp. Educ. 1945, 14, 101 - 134.

LENTZ, T.F.: Acquiescence as a factor in the measurement of personality. Psychol. Bull. 1938, 35, 659.

LIENERT, G.A.: Testaufbau und Testanalyse. Weinheim 1967$^2$, 227 f.

MEDLEY, D.L. & MITZEL, H.E.: Verhalten im Unterricht, seine Erfassung durch Beobachtungsverfahren. In: GAGE, N.L. (Hrsg.): Handbuch der Unterrichtsforschung, Teil I, Weinheim, Berlin, Basel 1970, 770 - 789.

MORSH, J.E.: Development report - Systematic observation of instructor behavior. In: USAF Pers. Train. Res. Cent. Develpm. Rept, 1956, No. AFPTRC - TN - 56 - 52.

MADSEN, C.H., WESLEY, W.C. & R.D. THOMAS: Rules, Praise, and Ignoring: Elements of Elementary Classroom Control. J. of Applied Analysis Behavior, Summer 68, 1, 139 - 150.

MOWRER, O.H.: Learning theory and the symbolic process. New York: Wiley 1960.

NIEMANN, H.J.: Entwicklung von Einstellungsskalen zum "Lehrer-Engagement" bei Lehrerinnen in der ersten Grundschulklasse. In: Zt. für Entwicklungspsychol. und Pädagogische Psychol. 1972, Bd. IV, Heft 2, 118 - 129.

PETER, G.: Soziale Interaktion in der Schulklasse. Deutsche Bearbeitung des Kapitels 13 von WITHALL, J. & W.W. LEWIS: Social Interaction in the classroom. In: GAGE, N.L. (Hrsg.): Handbuch der Unterrichtsforschung, Teil II, Weinheim, Berlin, Basel 1970.

RASCH, G.: Probabilistic models for some intelligence and attainment tests. Kopenhagen 1960.

RASCH, G.: An informal report on a theory of objectivity in comparisons. Proceedings of the NUFFIC International summer session in science at "Het Onde Hof", The Hague, July 14. - 18., 1966 a.

RASCH, G.: An individual - centered approach to item - analysis with two categories of answers. Proceedings of the NUFFIC International summer session in science at "Het Onde Hof", The Hague, July 14. - 18., 1966 b.

REMMERS, H.H.: Schätzverfahren in der Unterrichtsforschung. in: GAGE, N.L. (Hrsg.): Handbuch der Unterrichtsforschung, Teil I, Weinheim, Berlin, Basel 1970, 962 - 963.

RORER, L.G.: The great response - style myth. Psychol. Bull. 1965, 63, 129 - 156.

SACHS, L.: Statistische Auswertungsmethoden. Berlin, Heidelberg, New York 1969.

SCHAEFER, E.S. & R.Q. BELL: Development of a parental attitude research instrument. Child Develp. 1968, 29, 339 - 361.

SHOBEN, E.J.: The assessment of parental attitudes in relation to child adjustment. Genet. Psychol. Monogr. 1949, 39, 101 - 148.

SEGORD, P.F. & C.W. BACKMAN: Social Psychology. New York 1964.

SIEGELMAN, M.: Evaluation of Bronfenbrenner's questionnaire for children concerning parental behavior. Child Develp. 1965, 36, 163 - 174.

SIEGELMAN, M.: Loving and punishing parental behaviors in introversion tendencies in sons. Child Develp. 1966, 37, 985 - 992.

STAPF, K.H., HERRMANN, T., STAPF, A. & K.H. STÄCKER: Psychologie des elterlichen Erziehungsstils. Bern 1972.

STENE, J.: Einführung in Raschs Theorie psychologischer Messung. In: FISCHER, G.H. (Hrsg.): Psychologische Testtheorie. Bern 1968, 229 - 268.

SMITH, B.O.: A study of the logic of teaching: A report on the first phase of a five-year research project. Washington, DC: US Office of Education 1959.

TAUSCH, R. & A. TAUSCH: Erziehungspsychologie. Göttingen 1970.

TORGERSON, W.S.: Theory and Methods of Scaling. New York, London, Sydney 1958 7.

VELDMAN, D.J.: Fortran Programming for the Behavioral Sciences. Holt, Rinehart & Winston, New York 1968, 308 - 317.

WARD, J.H.: Hierarchical grouping to optimize an objective function. J. Am. Statist. Assoc. 58, 236 - 244.

WEBER, Erna: Grundriß der biologischen Statistik. Stuttgart 1967 6.

WETZEL, C.H.: Behavior Modification Techniques and the Training of Teacher's Aide. Psychol. in the School, 1970, 7, 325 - 330.

WHITMAN, M. & J. WHITMAN: Behavior Modification in the Classroom. Psychol. in the School 1971, 8, 176 - 186.

WIGGINS, J.S.: Social desirability in the MPPI and Welch' factor scales A and R.J. consult. Psychol. 1959, 23, 100 - 106.

WINER, B.J.: Statistical Principles in Experimental Design. London, New York 1970, 303 f.

WITHALL, J.: The development of a technique for the measurement of social-emotional climate in classrooms. J. exp. Educ. 1949, 17, 347 - 361.

WRIGHT, E.J. & V.H. PROCTOR: Systematic observation of verbal interaction as a method of comparing mathematic lessons. St. Louis, Mo: Washington Univ. (U.S. Office of Educ. Coop. Res. Project No. 816, 1961.

# A N H A N G

Name: ..... Alter: .....
Klasse: ..... Junge Mädchen

Anweisung

Liebe Schülerin, lieber Schüler,

auf den folgenden Seiten stehen einige Sätze und Fragen. Diese Sätze handeln von Deiner Lehrerin. Zum Beispiel steht der Satz da:

Die Lehrerin hat gute Laune
oft manchmal selten nie

Es gibt Lehrerinnen, die oft, manchmal, selten oder nie gute Laune haben. Wie ist es mit Eurer Lehrerin? Wenn für Deine Lehrerin das oft oder manchmal oder selten oder nie zutrifft, machst Du einen Kreis um das entsprechende Wort.
Bitte versuche die folgenden Fragen so zu beantworten, wie es nach Deiner Meinung für Deine Lehrerin ..... zutrifft.

Wir wollen mit dieser Befragung erfahren, wie Schüler über ihre Lehrer denken. Diese Befragung wird nicht nur bei Euch, sondern in vielen Klassen und Schulen durchgeführt. Diese Befragung wird von uns vertraulich behandelt. Eure Lehrerin erfährt nicht, wie der einzelne Schüler sie beurteilt hat.
Überlege bitte bei den einzelnen Sätzen nicht zu lange und achte darauf, daß Du keine Frage ausläßt.
Dies ist also keine Klassenarbeit und Du brauchst Dich nicht besonders zu beeilen.
Verschiedene Kinder brauchen zur Beantwortung verschieden lange.

Bitte nicht voneinander abgucken!

1. Der Lehrer freut sich, wenn er die Klasse betritt.

   oft manchmal selten nie

---

2. Wenn er etwas verspricht, dann hält er das auch.

   immer manchmal selten nie

---

3. Er achtet auf Pünktlichkeit.

   immer manchmal selten nie

---

4. Er gibt Klassenarbeiten schnell zurück.

   oft manchmal selten nie

---

5. Wenn er etwas anordnet, dann begründet er das auch.

   immer manchmal selten nie

---

6. Er hat es gern, wenn wir still auf unseren Plätzen sitzen.

   immer manchmal selten nie

---

7. Der Lehrer hat schlechte Laune.

   oft manchmal selten nie

---

8. Der Lehrer erklärt schwierige Dinge und Aufgaben so, daß alle, auch die schwächeren Schüler der Klasse, es verstehen.

   immer oft manchmal selten

---

9. Der Unterricht ist langweilig, daß man einschläft.

   nie selten manchmal oft

---

10. Der Lehrer läßt Vieles durchgehen.

    oft manchmal selten nie

---

11. Sie achtet auf genaue Einhaltung von Pausenzeichen.

immer manchmal selten nie

---

12. Wenn die Lehrerin etwas erklärt hat, sagt sie, daß es wohl nun auch dem Letzten klar geworden sein müsse.

oft manchmal selten nie

---

13. Die Lehrerin sagt, daß sie Fragen von Schülern für dumm hält.

oft manchmal selten nie

---

14. Die Lehrerin blamiert einige Schüler vor der Klasse.

oft manchmal selten nie

---

15. Die Lehrerin sagt uns gegenüber manchmal Dinge, die wir zu ihr nicht sagen dürfen.

nie selten manchmal oft

---

16. Auch wenn die Lehrerin bei Schülern Fehler entdeckt, bleibt sie freundlich

oft manchmal selten nie

---

17. Die Lehrerin unterstützt Schüler, die in der Klasse nicht so beliebt sind.

oft manchmal selten nie

---

18. Die Lehrerin erklärt, wie sie unsere Klassenarbeiten beurteilt.

immer manchmal selten nie

---

19. Die Lehrerin bespricht mit uns Schülern die Auswahl der Unterrichtsthemen.

oft manchmal selten nie

---

20. Die Lehrerin erteilt Befehle.

oft manchmal selten nie

---

21. Die Lehrerin gibt uns Aufgaben, ohne zu sagen, wie wir sie ausführen sollen.

nie selten manchmal oft

---

22. Ich gehe gern zur Schule.

immer manchmal selten nie

---

23. Die Lehrerin lobt uns, wenn wir gute Arbeit leisten.

immer manchmal selten nie

---

24. Wenn Schüler untereinander Streit haben, dann sagt die Lehrerin, wie die Schüler sich verhalten sollen.

immer manchmal selten nie

---

25. Die Lehrerin bringt uns Dinge bei, die wir gern können möchten.

immer manchmal selten nie

---

26. Wenn wir nicht sofort tun, was die Lehrerin sagt, wird sie ärgerlich.

oft manchmal selten nie

---

27. Wenn wir etwas getan haben, was die Lehrerin falsch findet, dürfen wir sagen, warum wir es getan haben.

immer manchmal selten nie

---

28. Man kann mit ihr über alles sprechen, auch wenn man etwas ausgefressen hat.

immer manchmal selten nie

---

29. Wenn die Lehrerin bestraft, merkt man, daß sie es nicht gern tut.

immer manchmal selten nie

---

30. Die Lehrerin sieht ein, daß die Schüler recht haben und gibt nach.

immer manchmal selten nie

---

31. Die Lehrerin schimpft mit uns.

oft manchmal selten nie

---

32. Wenn man viele Fragen hat, dann beantwortet sie die Lehrerin alle.

immer manchmal selten nie

---

33. Wenn in der Klasse Unruhe ist, dann bestraft die Lehrerin die ganze Klasse.

oft manchmal selten nie

---

34. Wenn man zu spät zur Schule kommt, dann hat die Lehrerin dafür Verständnis.

immer manchmal selten nie

---

35. Wenn man ein Schulbuch vergißt, dann schimpft die Lehrerin.

oft manchmal selten nie

---

36. Auch wenn man mal frech gewesen ist, bleibt die Lehrerin ruhig.

immer manchmal selten nie

---

37. Die Lehrerin sorgt dafür, daß alle Schüler arbeiten, so gut sie können.

immer manchmal selten nie

---

38. Bei unserer Lehrerin können wir uns viel herausnehmen.

oft manchmal selten nie

---

39. Die Lehrerin kann mich ganz gut leiden.

niemals selten manchmal meistens

---

40. Die Lehrerin weiß normalerweise, was sie will.

meistens manchmal selten nie

---

41. Die Lehrerin ist zu mir gerecht.

immer manchmal selten nie

---

42. Die Lehrerin gibt Strafarbeiten auf.

oft manchmal selten nie

---

43. Zu mir ist die Lehrerin höflich und nett.

meistens manchmal selten nie

---

44. Bei unserer Lehrerin dürfen wir offen unsere Meinung sagen.

immer meistens selten nie

---

45. An mir hat die Lehrerin etwas auszusetzen.

meistens manchmal selten nie

---

46. Die Lehrerin ist auch zu den Schülern nett, die faul sind.

immer manchmal selten nie

---

47. Die Lehrerin wird leicht ärgerlich.

oft manchmal selten nie

---

48. Die Lehrerin ist streng.

oft manchmal selten nie

---

Name:.. ..... .  Alter:.....

Vorname:..... ..  Junge  Mädchen

Klasse:... .....

Anleitung: Bitte beantworte jeden Satz, indem Du einen Kreis um das Oft bzw. Immer oder Manchmal oder Selten oder Nie machst. Es gibt keine richtigen und falschen Antworten. Arbeite schnell und denke nicht zu lange über den Sinn der Sätze nach. Es ist wichtig, daß Du keinen Satz ausläßt.

1. Ich komme zu spät, wenn ich pünktlich sein sollte.

oft  manchmal  selten  nie

2. Ich würde an der Grenze beim Zoll alles angeben, auch wenn ich wüßte, daß ich nicht kontrolliert werde.

immer  manchmal  selten  nie

3. Ich sage die Wahrheit.

immer  manchmal  selten  nie

4. Mein Benehmen ist gut und ordentlich.

immer  manchmal  selten  nie

5. Vor dem Essen wasche ich mir die Hände.

immer  manchmal  selten  nie

6. Ich beantworte einen Brief sobald wie möglich.

immer  manchmal  selten  nie

7. Ich bin pünktlich.

immer  manchmal  selten  nie

8. Ich mogele beim Spielen und beim Sport, um zu gewinnen.

oft  manchmal  selten  nie

9. Ich mache meine Hausaufgaben nur, wenn ich Lust dazu habe.

oft manchmal selten nie

---

10 Ich komme zu spät zur Schule oder zu einer Verabredung.

oft manchmal selten nie

---

11. Ich kann mich gut beherrschen.

immer manchmal selten nie

---

12. Ich bin in der Schule und zu Hause fleißig.

immer manchmal selten nie

---

13. Ich rede schlechte Dinge über andere.

oft manchmal selten nie

---

14. Leuten, die ich nicht mag, wünsche ich Schlechtes.

oft manchmal selten nie

---

15. Ich esse alles, was es zu Hause gibt.

immer manchmal selten nie

---

16. Ich gebe vor anderen Menschen an, um einen besseren Eindruck zu machen.

oft manchmal selten nie

---

17. Ich habe ein Versprechen gebrochen, das ich hätte halten sollen.

oft manchmal selten nie

---

18. Ich spreche über Dinge, von denen ich nichts verstehe.

oft manchmal selten nie

---

19. Es fallen mir Gedanken ein, die andere Menschen nicht wissen dürfen.

oft manchmal selten nie

---

20. Ich lüge, um mir Vorteile zu verschaffen.

oft manchmal selten nie

---

21. Es gibt Unterrichtsfächer in der Schule, die ich nicht mag.

meistens manchmal selten nie

---

22. Ich verschiebe gern etwas auf morgen, was ich heute hätte tun sollen.

oft manchmal selten nie

---

23. Ich tue sofort, was man mir sagt.

immer manchmal selten nie

---

24. Ich bin ruhig und geduldig.

immer manchmal selten nie

---

25. Ich mache meine Arbeit sofort und sobald wie möglich.

immer manchmal selten nie

---

26. Ich bin ein guter Verlierer beim Spielen und beim Sport.

immer manchmal selten nie

---

Instruktion zum Ausfüllen des Polaritätenprofils

Im folgenden werden Ihnen 25 Eigenschaftswörter und deren Gegensätze (z.B. freundlich - unfreundlich) vorgelegt, mit deren Hilfe Sie bitte jede(n) Schüler (in) Ihrer Klasse charakterisieren wollen.
Zwischen den Eigenschaftspaaren befindet sich eine 7 - Punkte - Skala. Mit Hilfe dieser Skala können Sie die Ausprägung einer der beiden Eigenschaften bei dem zu beurteilenden Schüler angeben.
Bei der Beurteilung rufen Sie sich bitte erst den betreffenden Schüler (z.B. Franz Maier) ins Gedächtnis.
Sodann überlegen Sie bitte, welche der beiden Eigenschaften (z.B. aufmerksam oder abgelenkt) den Schüler besser beschreibt. Danach bestimmen Sie bitte noch die Ausprägung dieser Eigenschaften.
Dabei bedeutet:
Der Punktwert 3 eine starke, der Punktwert 2 eine mittlere und der Punktwert 1 eine schwache Ausprägung. Der Punktwert 0 bedeutet, daß keine der beiden Eigenschaften zutrifft.
Halten Sie beispielsweise den Schüler Franz Maier eher für aufmerksam als für abgelenkt, dann durchkreuzen Sie bitte eine der drei Klammern auf der linken Seite der Skala und zwar:
Ist der Schüler sehr aufmerksam, dann durchkreuzen Sie die Klammer mit der 3, bei mittlerer Aufmerksamkeit die Klammer mit der 2, bei schwacher Aufmerksamkeit die Klammer mit der 1 (jeweils auf der linken Seite der Skala).
Halten Sie den Schüler eher für abgelenkt, dann durchkreuzen Sie bitte in analoger Weise eine der drei Klammern auf der rechten Seite, halten Sie den Schüler weder für aufmerksam noch für abgelenkt, dann durchkreuzen Sie bitte die Klammer mit der 0.
Nun noch eines. Bei der Beurteilung der Schüler durch solche Polaritätenprofile kann man die Einschätzung aufgrund sehr unterschiedlich langer und sorgfältiger Überlegungen abgeben.

Außerdem wechselt das Verhalten der Schüler in unterschiedlichen Situationen und damit auch die Beurteilung. Stellen Sie sich bitte das typische (häufigste) Verhalten der Schüler im Klassenverband vor und geben Sie bitte dann ruhig ein intuitives Urteil ab.
Das heißt, überlegen Sie bitte nicht zu lange und geben Sie bei der Beurteilung dem ersten Eindruck nach.

Name und Vorname des Schülers: FRANZ MAIER

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1. | uninteressiert | (3) | (2) | (1) | (0) | (1) | (2) | (3) | interessiert |
| 2. | verschlossen | (3) | (2) | (1) | (0) | (1) | (2) | (3) | offen |
| 3. | sensibel | (3) | (2) | (1) | (0) | (1) | (2) | (3) | robust |
| 4. | selbstsicher | (3) | (2) | (1) | (0) | (1) | (2) | (3) | selbstunsicher |
| 5. | unbegabt | (3) | (2) | (1) | (0) | (1) | (2) | (1) | begabt |
| 6. | einfallsreich | (3) | (2) | (1) | (0) | (1) | (2) | (3) | einfallslos |
| 7. | gesellig | (3) | (2) | (1) | (0) | (1) | (2) | (3) | ungesellig |
| 8. | geltungsbedürftig | (3) | (2) | (1) | (0) | (1) | (2) | (3) | bescheiden |
| 9. | fleißig | (3) | (2) | (1) | (0) | (1) | (2) | (3) | faul |
| 10. | kompliziert | (3) | (2) | (1) | (0) | (1) | (2) | (3) | unkompliziert |
| 11. | intelligent | (3) | (2) | (1) | (0) | (1) | (2) | (3) | dumm |
| 12. | ehrgeizig | (3) | (2) | (1) | (0) | (1) | (2) | (3) | gleichgültig |
| 13. | unkonzentriert | (3) | (2) | (1) | (0) | (1) | (2) | (3) | konzentriert |
| 14. | höflich | (3) | (2) | (1) | (0) | (1) | (2) | (3) | frech |
| 15. | unordentlich | (3) | (2) | (1) | (0) | (1) | (2) | (3) | ordentlich |
| 16. | schüchtern | (3) | (2) | (1) | (0) | (1) | (2) | (3) | dreist |
| 17. | folgsam | (3) | (2) | (1) | (0) | (1) | (2) | (3) | unfolgsam |
| 18. | Unruhig | (3) | (2) | (1) | (0) | (1) | (2) | (3) | ruhig |
| 19. | pflichtbewußt | (3) | (2) | (1) | (0) | (1) | (2) | (3) | pflichtvergesse |
| 20. | aufmerksam | (3) | (2) | (1) | (0) | (1) | (2) | (3) | abgelenkt |
| 21. | aufgeweckt | (3) | (2) | (1) | (0) | (1) | (2) | (3) | träge |
| 22. | ausgelichen | (3) | (2) | (1) | (0) | (1) | (2) | (3) | launisch |
| 23. | zuverlässig | (3) | (2) | (1) | (0) | (1) | (2) | (3) | unzuverlässig |
| 24. | sympatisch | (3) | (2) | (1) | (0) | (1) | (2) | (3) | unsympatisch |
| 25. | führend | (3) | (2) | (1) | (0) | (1) | (2) | (3) | zurückhaltend |

Lebenslauf

Am 19. August 1938 wurde ich als Sohn des kaufmännischen Angestellten Fritz Masendorf und seiner Ehefrau Erna, geb. Degener, in Osterode am Harz geboren.

Von 1946 bis 1952 besuchte ich die Volksschule, von 1952 bis 1956 den Aufbauzug der Volksschule in Osterode. Das Abitur legte ich im März 1960 am Gymnasium in Osterode ab. Anschließend begann ich mit dem Lehrerstudium, das ich im Februar 1963 an der Pädagogischen Hochschule in Braunschweig mit der Ersten Lehrerprüfung für das Lehramt an Volksschulen abschloß.

Vom Sommersemester 1963 an studierte ich an der Universität Münster drei Semester Psychologie, danach war ich von Januar 1965 bis Mitte Oktober 1966 als Hauptschullehrer in der Freien Hansestadt Bremen tätig.
Im Wintersemester 1966/67 setzte ich das Psychologiestudium von insgesamt 10 Semestern an der Universität Marburg fort. Das Vordiplom legte ich im Oktober 1967, das Hauptdiplom im April 1970 ab.

Vom 1. 5. 1970 bis 30. 11. 1971 war ich an der Pädagoggischen Hochschule Ruhr in Dortmund tätig, zunächst als Verwalter einer wissenschaftlichen Assistentenstelle am Lehrstuhl für Lernbehindertenpädagogik an der Abteilung Heilpädagogik, ab 1. 3. 1971 als Empiriker in der Forschungsgruppe zur wissenschaftlichen Begleitung des Gesamtschulversuchs NRW.

Seit dem 1. 12. 1971 bin ich wissenschaftlicher Assistent im Fachbereich Psychologie an der Pädagogischen Hochschule Niedersachsen, Abteilung Osnabrück.

Im August 1964 heiratete ich die Diplom-Bibliothekarin Elke Masendorf, geb. Schwartz. Meine Tochter Gesine wurde im März 1965 geboren.